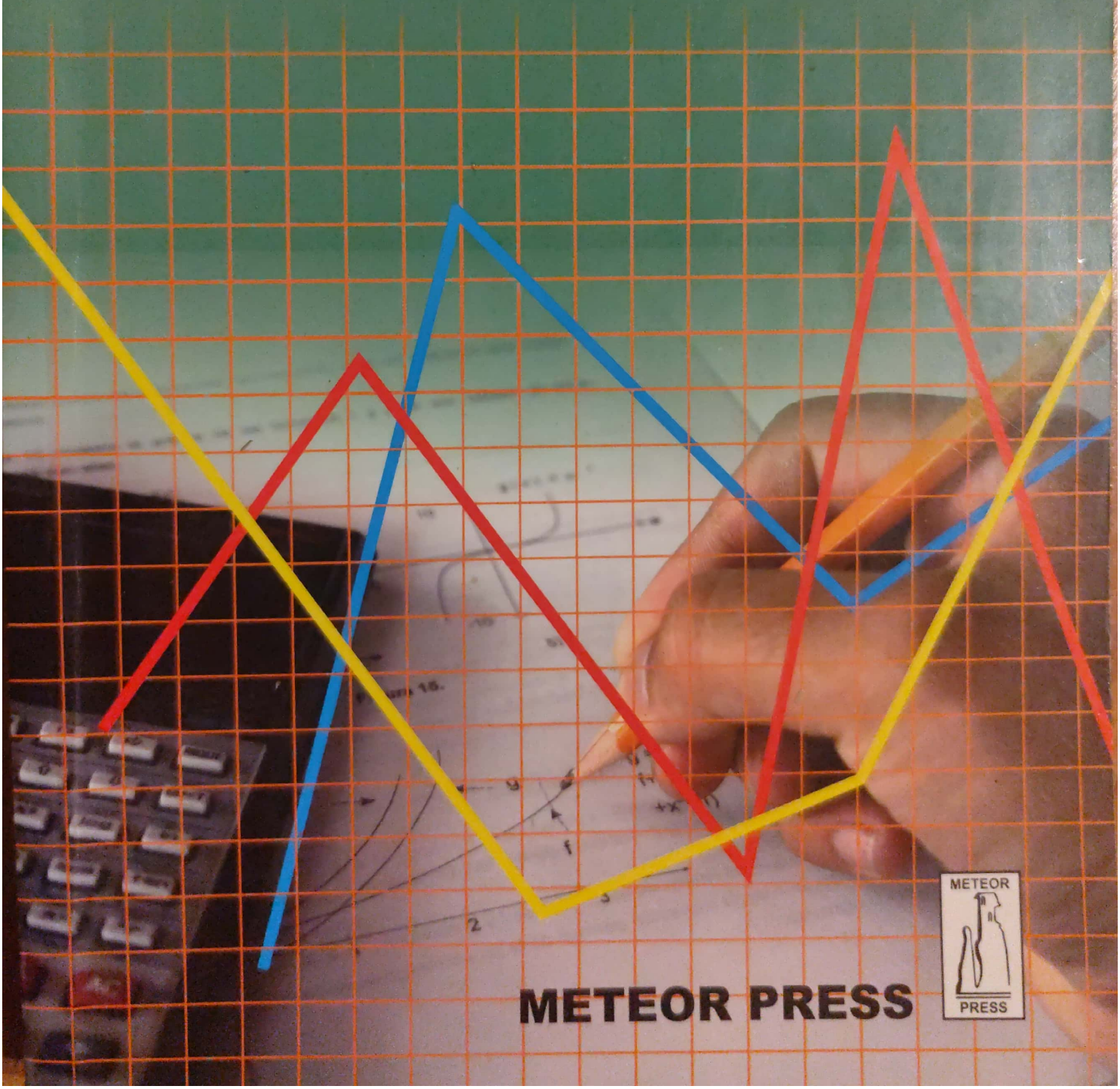
Simona Ioana Ghiță

STATISTICĂ



METEOR PRESS

Scanned with OKEN Scanner

Redactor: *Anca Lepădatu*
Tehnoredactor: *Rodica Boacă*
Copertă: *Mihai Chiriacescu*

Meteor Press, editură acreditată de C.N.C.S.I.S.

cu nr. 1142/iunie 2003

www.meteorpress.ro

© 2006 Toate drepturile asupra acestei ediţii sunt rezervate editurii **METEOR PRESS.**

Str. Bahluiului nr. 1, 011281, sector 1, Bucureşti
C.P. 41-128
Tel.: (021)222.33.12; Fax: (021)222.83.80
E-mail: editura@meteorpress.ro

Distribuţie la:
Tel./Fax: (021)222.33.12; (021)222.83.80
E-mail: carte@meteorpress.ro

ISBN: 978-973-728-152-4
ISBN: 973-728-152-7

Tiparul executat la ***S.C. LUMINA TIPO s.r.l.***
str. Luigi Galvani nr. 20 bis, sect. 2, Bucureşti, tel./fax 211.32.60; tel. 212.29.27
E-mail: **office@luminatipo.com**, *www.luminatipo.com*

Scanned with OKEN Scanner

SIMONA IOANA GHIȚĂ

STATISTICĂ

METEOR PRESS

Scanned with OKEN Scanner

CUPRINS

Capitolul 1. INTRODUCERE ÎN STUDIUL STATISTICII 11
1.1. Ce este statistica? De ce studiem statistica? 11
1.2. Ce studiază statistica? 12
1.3. Momente ale evoluției istorice a statisticii 13
1.4. Concepte de bază utilizate în statistică 15
1.5. Scale de măsurare a datelor statistice 21
1.6. Procesul de investigație statistică și etapele sale 25
1.7. Organizarea statisticii oficiale în România și pe plan internațional 27

Capitolul 2. CULEGEREA ȘI PRELUCRAREA PRIMARĂ A DATELOR STATISTICE 34
2.1. Observarea statistică 34
2.1.1. Importanța și conținutul observării statistice 34
2.1.2. Programul observării statistice 35
2.1.3. Metode de observare statistică 39
2.1.4. Conceptul de „eroare statistică". Modalități de control al datelor statistice 46
2.2. Prelucrarea primară a datelor statistice 49
2.2.1. Sistematizarea datelor statistice: clasificarea și gruparea datelor statistice 50
2.2.2. Modalități de prezentare și reprezentare a datelor statistice 59
2.2.2.1. Serii statistice 59
2.2.2.2. Grafice și tabele statistice 71

Capitolul 3. ANALIZA DESCRIPTIVĂ A SERIILOR STATISTICE UNIVARIATE 89
3.1. Principalele aspecte ale analizei statistice a seriilor univariate 89
3.2. Caracterizarea numerică a tendinței centrale 89
3.2.1. Determinarea mediei, medianei și a modului pentru o serie de date negrupate 90
3.2.2. Determinarea mediei, medianei și a modului pentru o distribuție de frecvențe pe variante 94

Scanned with OKEN Scanner

3.2.3. Determinarea mediei, medianei şi a modului pentru o distribuţie de frecvenţe pe intervale de variaţie 98
3.2.4. Proprietăţile mediei aritmetice 102
3.2.5. Alţi indicatori de poziţie 113
3.2.6. Analiza comparativă a principalilor indicatori ai tendinţei centrale 116
3.2.7. Medii cu aplicaţii speciale utilizate în statistică 120
3.3. Analiza statistică a variabilităţii 125
3.3.1. Indicatorii simpli de măsurare a variabilităţii 127
3.3.2. Indicatorii sintetici de măsurare a variabilităţii 130
3.3.3. Măsurarea dispersării valorilor unei caracteristici într-o populaţie statistică structurată pe grupe 144
3.4. Analiza formei distribuţiilor de frecvenţe 151
3.4.1. Boltirea (excesul) 151
3.4.2. Asimetria (oblicitatea) 154

Capitolul 4. UTILIZAREA SONDAJULUI STATISTIC ÎN STUDIUL FENOMENELOR ECONOMICO-SOCIALE 158
4.1. Introducere. Definiţie, avantaje şi dezavantaje ale sondajului statistic .. 158
4.2. Noţiuni specifice sondajului statistic 159
4.3. Modalităţi de extracţie aleatoriu a eşantioanelor. Tipuri de sondaj 163
4.4. Posibilităţi de estimare a parametrilor colectivităţii generale 167
4.5. Distribuţia de eşantionare a mediei de selecţie 168
4.6. Sondajul simplu aleatoriu pentru o caracteristică nealternativă 173
4.6.1. Determinarea erorilor de sondaj şi a intervalului de încredere 173
4.6.2. Dimensionarea eşantionului în cazul sondajului simplu aleatoriu 180
4.7. Sondajul aleatoriu simplu pentru o caracteristică alternativă 182
4.7.1. Determinarea erorilor şi a intervalului de încredere pentru proporţii 182
4.7.2. Dimensionarea eşantionului în cazul unei variabile alternative ... 185
4.8. Determinarea erorilor şi a intervalului de încredere în cazul eşantioanelor de volum redus 186
4.9. Sondajul aleatoriu tipic (stratificat) 190
4.9.1. Noţiuni specifice pereche utilizate în cadrul sondajului tipic 190
4.9.2. Determinarea erorilor, a intervalului de încredere şi a volumului eşantionului în cazul sondajului aleatoriu tipic 193
4.10. Elemente de testare a ipotezelor statistice 197

Scanned with OKEN Scanner

Capitolul 5. ANALIZA STATISTICĂ A SERIILOR DE DATE MULTIVARIATE 205
5.1. Conceptul de legătură statistică. Clasificarea legăturilor statistice 205
5.2. Metode elementare de analiză a legăturilor dintre variabilele statistice ... 208
5.3. Metode parametrice de analiză a legăturilor dintre variabilele statistic . 212
5.3.1. Metoda regresiei liniare simple 213
5.3.1.1. Determinarea parametrilor funcției de regresie liniară ... 215
5.3.1.2. Testarea semnificației parametrilor de regresie și a validității modelului de regresie liniară simplă 221
5.3.1.3. Previzionarea evoluției variabilei rezultative pe baza modelului de regresie liniară simplă 227
5.3.1.4. Analiza statistică a factorilor aleatorii 230
5.3.2. Metoda corelației liniare simple 234
5.3.3. Regresia și corelația neliniară simplă 241
5.3.4. Regresia și corelația liniară multiplă 244
5.4. Elemente de analiză dispersională (ANOVA) 248
5.5. Metode neparametrice utilizate în analiza legăturilor dintre variabilele statistice 257
5.5.1. Asocierea variabilelor nominale 257
5.5.2. Asocierea variabilelor ordinale. Metoda corelației rangurilor 262

Capitolul 6. ANALIZA EVOLUȚIEI ÎN TIMP A FENOMENELOR ECONOMICO-SOCIALE 266
6.1. Particularitățile și tipologia seriilor cronologice 266
6.2. Indicatorii seriilor cronologice 268
6.2.1. Indicatori absoluți 269
6.2.2. Indicatori relativi 271
6.2.3. Indicatori medii 274
6.3. Componentele unei serii cronologice 278
6.4. Metode de determinare a tendinței de lungă durată 283
6.4.1. Metode mecanice de determinare a trendului unei serii cronologice 283
6.4.2. Modele analitice de determinare a trendului unei serii cronologice 292
6.4.3. Metode de alegere a celui mai bun model de determinare a tendinței 297
6.5. Modelarea componentei sezoniere a unei serii cronologice 298
6.6. Analiza componentei ciclice a unei serii cronologice 305

Scanned with OKEN Scanner

6.7. Analiza componentei aleatorii a unei serii cronologice 306
6.8. Covariația. Covariație cu întârziere (decalaj) 307
6.9. Previzionarea pe baza seriilor cronologice 308

Capitolul 7. INDICII STATISTICI 313
7.1. Mărimi relative. Conținut, formă generală, tipologie 313
7.2. Indicii ca mărimi relative. Rolul și funcțiile indicilor în cercetarea statistică 322
7.3. Clasificarea indicilor 323
7.4. Reguli de construire a indicilor elementari și sintetici 324
7.4.1. Indici individuali (elementari) 324
7.4.2. Indici de grup (sintetici) 326
7.4.2.1. Indici sintetici calculați ca indici agregați 326
7.4.2.2. Indici sintetici calculați ca medie a indicilor individuali ... 330
7.4.2.3. Indici de grup calculați ca raport a două medii 333
7.5. Sisteme de ponderare folosite în construirea indicilor sintetici 338
7.6. Determinarea contribuției factorilor la variația unui fenomen complex prin metoda indicilor 341
7.6.1. Metoda substituției în lanț 341
7.6.2. Metoda influențelor izolate ale factorilor (metoda restului nedescompus) 347
7.7. Serii de indici statistici 352
7.8. Sisteme concrete de indici folosite în activitatea economico-socială 356
7.9. Indici teritoriali 363

ANEXE 367

BIBLIOGRAFIE 376

Scanned with OKEN Scanner

Cuvânt-înainte

Statistica este un „numitor comun" al tuturor ştiinţelor, un limbaj comun al acestora, este un set de instrumente aplicabil în toate domeniile investigaţiei ştiinţifice, dar şi arta mânuirii corecte a acestora. Învăţându-ne să raţionăm corect, statistica ne ajută, astfel, să luăm deciziile optime în orice aspect al realităţii înconjurătoare.

Lucrarea de faţă îşi propune nu numai să treacă în revistă, dar şi să aprofundeze într-un mod complet, logic, unitar şi accesibil, bogăţia metodologiei statistice; astfel, lucrarea ne introduce în universul vocabularului statistic, ne prezintă principalele etape ale demersului statistic, diversele metode statistice ce pot fi utilizate în rezolvarea unor situaţii reale, practice, ne explică formulele şi conţinutul indicatorilor statistici, ne arată cum putem „traduce", interpreta şi folosi rezultatele prelucrării statistice în vederea adoptării unor decizii eficiente. Aspectul teoretic al prezentării metodelor statistice este dublat de ilustrarea modului de aplicabilitate a lor, printr-o gamă largă de exemple din numeroase domenii ale vieţii economico-sociale. Unele din aceste exemple prin care se identifică probleme reale, din viaţa de zi cu zi, sunt rezolvate cu ajutorul programelor informatice specializate în prelucrarea şi analiza datelor statistice.

Prezenta lucrare se adresează studenţilor din învăţământul superior economic, dar şi celor din alte specializări ale învăţământului universitar, ce folosesc instrumentarul statistic în investigarea fenomenelor de masă din cele mai variate domenii. Ea poate fi utilă, totodată şi cercetătorilor, specialiştilor sau tuturor celor interesaţi în aplicarea vastei metodologii pe care statistica o pune la dispoziţie pentru întreaga cercetare ştiinţifică.

Autoarea mulţumeşte tuturor celor care au avut sau vor avea amabilitatea ca, prin sugestii şi sfaturi competente, să contribuie la completarea şi îmbunătăţirea prezentei lucrări.

Autoarea

Scanned with OKEN Scanner

CAPITOLUL 1
INTRODUCERE ÎN STUDIUL STATISTICII

1.1. Ce este statistica? De ce studiem statistica?

Statistica este un „numitor comun" al tuturor ştiinţelor, un limbaj comun al acestora, un set de instrumente ce poate fi aplicat în toate domeniile cercetării, dar şi „arta" de mânuire corectă a acestor instrumente. În viaţa de zi cu zi, statistica ne oferă o paletă largă de tehnici cu aplicabilitate în cele mai variate câmpuri de interes care ne ajută să rezolvăm problemele practice şi ne învaţă să raţionăm corect. Pentru unii, statistica înseamnă numărul de goluri, de lovituri libere sau de cornere înscrise într-un meci de fotbal, pentru alţii înseamnă nivelul Produsului Intern Brut, numărul şomerilor, rata de ocupare a forţei de muncă, indicele preţurilor sau rata inflaţiei, alături de alte asemenea date regăsite într-o publicaţie de specialitate. Pentru alţii statistica poate însemna o ramură a ştiinţei, o ştiinţă de sine stătătoare sau o disciplină de învăţământ. „Intrumentarul" aceste ştiinţe se poate aplica, practic, în toate celelalte ştiinţe, fie că este vorba de psihologie, de medicină, fizică, chimie, meteorologie sau istorie.

Aşadar, *statistica reprezintă ştiinţa colectării datelor, a prezentării lor într-o formă sistematică, a analizării acestora şi a interpretării informaţiilor numerice obţinute*. Pe baza acestei succesiuni de operaţii, statistica ajută la luarea deciziilor. Rolul ei nu este, aşadar, acela de a emite decizii, ci de a orienta decidentul pe o bază fundamentată asupra deciziei optime pe care o va adopta.

Necesitatea cuantificării, a exprimării numerice a fenomenelor şi proceselor se manifestă în toate domeniile vieţii economico-sociale şi acest lucru îl face şi statistica, dar nu numai atât. Ea permite caracterizarea, prin doar câteva nivele numerice, a unor largi grupuri de date, fără a fi necesar să investigăm fiecare caz în parte. Acest lucru asigură o economie de resurse materiale, financiare, umane şi temporale. Mai mult, în urma investigării şi prelucrării setului de date există posibilitatea efectuării unor inferenţe statistice.

Etimologia cuvântului *statistică* indică originea latină, de la termenul *status*, ce desemnează o stare, o situaţie, un fapt. Astăzi, *statistica reprezintă un ansamblu de principii şi metode cu ajutorul cărora putem obţine informaţii utile despre un set de date*.

Scanned with OKEN Scanner

Domeniul statisticii poate fi subdivizat în două arii: ***statistica descriptivă*** şi ***statistica inferenţială***. De obicei, ***statistica descriptivă*** este sensul mai cunoscut, mai uzual al colectării, organizării, sintetizării, prezentării şi descrierii datelor numerice şi nenumerice, într-o formă convenabilă, în timp ce termenul de ***inferenţă*** se referă la tehnicile de interpretare a rezultatelor obţinute prin metodele statisticii descriptive, utilizate apoi la luarea deciziilor. Deseori, decidenţii sunt nevoiţi să adopte decizii pe baza datelor care nu provin de la toate unităţile colectivităţii investigate, ci de la o parte a acestora şi, cu ajutorul statisticii inferenţiale, ei pot formula concluzii valabile, în anumiţi termeni probabilistici, pentru întreaga colectivitate.

De exemplu, metodele statisticii descriptive pot ajuta pe managerul unei reţele de magazine de produse electrocasnice să compare vânzările săptămânale din ultimul an în două puncte de vânzare diferite, astfel: sunt sintetizate vânzările săptămânale (eventual grupate pe tipuri de produse electrocasnice) în câteva nivele numerice: nivelul mediu al vânzărilor săptămânale, gradul de variaţie al vânzărilor faţă de media lor. Tabelele şi graficele ajută la prezentarea mai facilă, mai clară şi mai uşor de reţinut a informaţiilor obţinute, astfel încât să poată fi repede identificate diferenţele esenţiale dintre vânzările celor două puncte de vânzare.

De asemenea, managerul ar putea să dorească deschiderea a încă unui punct de vânzare într-o anumită zonă a unui oraş; el poate organiza o cercetare statistică selectivă în cadrul căreia să afle dacă cetăţenii incluşi într-un *eşantion* apreciază ca pozitivă această iniţiativă şi dacă ar fi interesaţi să-şi facă cumpărăturile în noul punct de vânzare. Managerul va extinde rezultatele cercetării eşantionului la nivelul întregii populaţii şi astfel spunem că el face o inferenţă statistică.

1.2. Ce studiază statistica?

Statistica studiază fenomenele şi procesele care prezintă următoarele particularităţi:

– se întâlnesc la un număr mare de unităţi;
– variază ca nivel de la o unitate la alta;
– sunt definite, circumscrise în timp, în spaţiu şi ca structură organizatorică (adică sunt entităţi distincte circumscrise în timp, spaţiu şi organizatoric).

Aceste fenomene se numesc ***fenomene de masă*** şi constituie obiectul de studiu al statisticii.

Pentru a putea cunoaşte aceste fenomene, nu este de ajuns să studiem câteva unităţi purtătoare a acestor fenomene, ci este necesar să avem viziunea

Scanned with OKEN Scanner

ansamblului, a întregului, să studiem un număr suficient de mare de cazuri pentru ca ele să fie sub incidența legii numerelor mari.

Fenomenele studiate de statistică (***fenomene stochastice*** sau ***statistice***) sunt fenomene *nedeterministe*, influențate, de regulă, de un număr mare de factori cu acțiune sistematică sau aleatorie, esențială sau neesențială, cu acțiune într-un același sens sau în sensuri diferite. Acest tip de fenomene sunt guvernate de **legi statistice**, care se manifestă ca tendință și care pot fi cunoscute, identificate și studiate doar prin cercetare ansamblului unităților la care se produc.

În antiteză cu fenomenele stochastice, există ***fenomene deterministe***, univoc determinate, care apar sub influența unui număr restrâns de cauze și care se manifestă cu aceeași valoare/nivel, dacă și cauzele care le-au generat sunt aceleași. Spre deosebire de fenomenele de masă, fenomenele deterministe pot fi cunoscute, identificate și cercetate pe cale experimentală. Totodată, valabilitatea lor poate fi probată la nivelul fiecărui caz individual în parte (așa se întâmplă, de regulă, cu fenomenele din științele naturii – fizică, chimie etc.).

1.3. Momente ale evoluției istorice a statisticii

Statistica a răspuns nevoii de cunoaștere, sub aspect cantitativ, a diferitelor aspecte ale vieții economico-sociale. Și cum această necesitate există dintotdeauna, și statistica a apărut sub forma unor înregistrări izolate, singulare, din cele mai vechi timpuri, încă din antichitate, aceste manifestări alcătuind așa numita ***statistică practică***. Statistica practică este veche de peste 4 milenii și îmbracă forma înregistrării numărului de nașteri sau căsătorii, în preajma templelor, evidența stocurilor, a rezervelor de hrană, a pământurilor sau a aurului deținute de unele personalități, a producției agricole etc., îndeosebi după apariția proprietății private și a formațiunilor sclavagiste. Mai mult, chinezii, grecii, egiptenii și romanii organizau recensăminte ale populației, încă din mileniile 5-4 î.Hr.

Pe măsura apariției statelor feudale, apar tot mai multe descrieri ale aspectelor geografico-naturale, politice, economico-sociale, istorice ale statelor. În aceste „studii", în care se folosesc și date statistice, se regăsesc rădăcinile teoretice ale statisticii, etapă numită ***statistica descriptivă***; întemeietorul școlii descriptive a fost Hermann Conring (1606-1681), profesor la Universitatea din Helmstedt, căruia îi aparține primul curs de statistică (statistica devine acum disciplină de învățământ). Un alt reprezentant al acestei școli a fost Gottfried Achenwall (1719-1772), considerat și „părintele statisticii", deoarece el utilizează, pentru prima dată într-o lucrare, cuvântul *statistică*, cu sensul de *stare*, *situație*. Studiile elaborate în această perioadă surprindeau în mod descriptiv diversele aspecte ale situației geografice,

Scanned with OKEN Scanner

economice, politice statale, însă ele nu cuprindeau şi o analiză cauzală a fenomenelor, de descoperire a legilor care le guvernau.

Acest ultim aspect se regăseşte, însă, în lucrările elaborate în Anglia, odată cu apariţia şi dezvoltarea societăţii capitaliste. Se conturează acum o altfel de modalitate de cunoaştere a fenomenelor economico-sociale, cercetarea experimentală, asemănătoare cu cea folosită în fizică, chimie, în ştiinţele naturale – în general. Aceasta a dus la apariţia statisticii ca ştiinţă, etapă numită ***aritmetica politică***, denumire ce provine de la lucrarea lui William Petty (1623-1687). Lucrările elaborate în această perioadă sunt caracterizate de o mai mare exactitate în cercetarea realităţilor economico-sociale, cu un mai pronunţat caracter ştiinţific, împrumutând din rigurozitatea metodelor matematice (acum se folosesc unele noţiuni de bază întâlnite în statistica de astăzi – medie, proporţie etc.). Deşi numele şcolii provine de la lucrarea lui William Petty, întemeietorul ei este considerat John Ground (1620-1674), autorul unor studii demografice asupra populaţiei din Londra, studii în care el a determinat anumite regularităţi în mişcarea naturală a populaţiei.

În cercetările statistice se pune un accent tot mai mare pe metodele exacte, cantitative, aşa cum se întâmplă în scrierile lui L.A.I. Quetelet (1796-1874), considerat ca fiind cel care a pus bazele statisticii moderne, întrucât, în studiile sale în domeniul demografiei şi criminalităţii, el a folosit elemente din teoria probabilităţii şi legea numerelor mari. Nume marcante şi-au adus contribuţia la dezvoltarea acestei noi şcoli statistice numite ***calculul probabilităţilor***, între care J. Bernoulli (cel care a pus bazele teoriei numerelor mari), K.F. Gauss şi P.S. Laplace, cei care au dezvoltat conceptul şi teoria legii normale de repartiţie, R.A. Fisher, G.U. Yulle, K. Pearson, Cebîşev, Liapunov, Markov şi alţii. Tot acum se elaborează şi principiile teoriei selecţiei, prin utilizarea combinată a metodei deductive cu cea inductivă.

Teoria matematică a comunicaţiei, elaborată de Cl. Shannon (1948) marchează începutul unei noi etape în evoluţia statisticii: ***etapa informaţională***, continuată apoi cu ***etapa sistemică***, în care statistica este privită prin prisma teoriei generale a sistemelor (L. von Bertalanffy, N. Wiener). Asupra statisticii şi-au pus, deci, amprenta, progresele înregistrate în alte ştiinţe (informatică, cibernetică etc.), dar, şi ea, la rândul ei, îşi pune aparatul metodologic în slujba altor ştiinţe, precum econometria, sociometria, managementul, marketingul sau cercetările operaţionale. Statistica are, aşadar, puncte comune cu toate aceste ştiinţe, însă ea se delimitează ca o ştiinţă de sine stătătoare, cu obiect şi metodologie proprie. Statistica este un instrument indispensabil în studierea şi cunoaşterea realităţilor economico-sociale într-o viziune sistemică, la nivel micro, mezo şi macroeconomic.

Scanned with OKEN Scanner

1.4. Concepte de bază utilizate în statistică

Pentru a efectua o cercetare statistică asupra unui fenomen şi pentru a comunica şi altora rezultatele şi concluziile studiului, sau pentru a înţelege concluziile la care au ajuns alţi specialişti în cercetările lor, trebuie să avem cu toţii un limbaj statistic comun. Vocabularul de bază al statisticii cuprinde următorii termeni:

1. **Populaţia statistică (colectivitatea statistică)**

Reprezintă totalitatea elementelor de acelaşi fel, cu trăsături observabile comune, studiate atunci când vrem să iniţiem un demers statistic. În ideea obţinerii tuturor informaţiilor de care avem nevoie şi a realizării unei economii de resurse materiale, financiare şi umane este foarte importantă definirea corectă a populaţiei statistice de interes, cu specificarea dimensiunii temporale, spaţiale şi organizatorice în care se înscrie populaţia. În trecut, prin populaţie se înţelegea o mulţime (o colectivitate) de persoane (indivizi), astăzi, însă, ea şi-a extins sensul, sfera de cuprindere, putându-se vorbi de populaţii de animale, obiecte, opinii, păreri, evenimente, măsurători etc. O populaţie se consideră integral definită, specificată dacă s-a elaborat lista completă a unităţilor care o compun. Unele populaţii statistice, având foarte multe elemente şi fiind, deci, foarte numeroase, pot fi considerate cu caracter infinit (de exemplu, populaţia tuturor absolvenţilor unei instituţii de învăţământ superior economic este infinită, deoarece ea poate include absolvenţii din trecut, prezent şi viitor); dacă facem însă nişte precizări, nişte delimitări, circumscrieri în timp, în spaţiu şi ca organizare, populaţiile infinite pot căpăta caracter finit (populaţia absovenţilor din Bucureşti, din anul universitar 2005-2006, de la Academia de Studii Economice).

2. **Unitatea statistică**

Este elementul, entitatea de sine stătătoare a unei populaţii statistice, care posedă o serie de trăsături caracteristice ce-i conferă apartenenţa la populaţia studiată. Se pot întâlni două tipuri de unităţi statistice:

– simple (un salariat, un agent economic, un obiect, o părere);

– complexe (o grupă de studenţi sau o echipă de salariaţi, o familie sau o gospodărie, o categorie de mărfuri).

Cum în societate fenomenele se întrepătrund, se poate întâmpla ca o aceeaşi unitate statistică să aparţină, simultan, mai multor populaţii: conform unor trăsături ea va aparţine unei colectivităţi, iar conform altor trăsături ea va aparţine altei colectivităţi. De exemplu, un student poate aparţine colectivităţii cunoscătorilor de limbă engleză (după limba străină cunoscută cel mai bine), sau colectivităţii

Scanned with OKEN Scanner

locuitorilor județului X (după localitatea de domiciliu), sau populației din sectorul de vârstă 15-24 ani (după vârstă) etc.

3. **Eșantionul (colectivitate de selecție, colectivitate parțială)**

Reprezintă un subset de elemente extrase (selectate) dintr-o populație totală, care, dacă este reprezentativ, întrunește aceleași caracteristici esențiale ca și populația totală. De cele mai multe ori, populațiile sunt prea numeroase pentru a fi studiate chiar toate elementele acestora, de aceea apelăm la eșantion, presupunând că, dacă el este reprezentativ pentru populația din care s-a extras (adică dacă el întrunește aceleași trăsături esențiale ca și populația totală), atunci indicatorii care descriu eșantionul pot fi aplicați și la nivelul populației totale.

4. **Inferența statistică**

Reprezintă o estimare, o previziune sau o decizie despre o populație, plecând de la informațiile asupra unui eșantion. De exemplu, pentru a calcula ratingul unei rețele de televiziune se estimează numărul potențialilor privitori care sunt conectați în rețea și cum numărul privitorilor este de ordinul milioanelor, inferența se bazează pe un eșantion de 1000 de persoane; sau putem lua un alt exemplu: un supervizor al calității producției estimează proporția produselor defecte realizate într-un proces de producție, utilizând un eșantion de câteva sute de unități – produse. Deoarece inferența statistică este realizată, de obicei, pe un subset relativ mic de elemente dintr-o populație de dimensiuni mai mari, metodele statistice nu pot realiza o estimație cu certitudine, ci doar cu o anumită probabilitate.

5. **Statistica (indicatorul statistic)**

Reprezintă o descriere numerică, cantitativă, obținută, de cele mai multe ori, pe baza datelor referitoare la un eșantion. Indicatorii statistici obținuți sunt variabile aleatorii, nu constante și se notează de obicei cu litere romane ($\bar{x}$, s, r etc.).

Orice indicator statistic are o *parte descriptivă* (noțională) și o parte *numerică*, un conținut real, o formulă de calcul și o anumită formă de exprimare. Indicatorii statistici pot fi elaborați în mod independent unii de alții, sau în interdependența acestora (în acest ultim caz, ei formează un sistem de indicatori). Dacă indicatorul statistic se obține direct din procesul de culegere a datelor sau printr-un proces de centralizare, atunci el se numește „indicator primar"; dacă însă indicatorul se obține prin aplicarea unui model de calcul, el se va numi „indicator derivat". De asemenea, dacă indicatorul caracterizează fiecare unitate statistică, el va fi un indicator analitic; dacă el se calculează și caracterizează o întreagă colectivitate statistică, atunci vom avea un indicator sintetic.

Scanned with OKEN Scanner

6. **Parametrul**

Este o valoare care sintetizează o caracteristică măsurabilă pentru o populaţie totală şi care este estimat prin intermediul statisticii. Parametrii populaţiei sunt constante şi se notează, de obicei, cu litere greceşti (μ, σ, δ etc.).

Statistica şi *parametrul* nu reprezintă, deci, acelaşi lucru: în timp ce statisticile sunt numere bazate pe observaţii directe şi măsurători, parametrii sunt valori estimate pe baza statisticilor calculate la nivelul eşantioanelor.

7. **Variabila statistică** (numită şi *caracteristică statistică* sau *variabilă aleatorie*)

Reprezintă o însuşire, o proprietate măsurabilă a unei unităţi statistice, întâlnită la toate unităţile care aparţin aceleiaşi colectivităţi şi care prezintă variabilitate (variază ca nivel) de la o unitate statistică la alta. În cadrul unei investigaţii statistice se înregistrează variabile de mai mare sau mai mică importanţă (esenţiale sau neesenţiale – în funcţie de scopul cercetării) de la unităţile populaţiei studiate (dacă efectuăm, de exemplu, un studiu privind nivelul de pregătire al studenţilor unei facultăţi, drept caracteristici esenţiale pot fi luate în considerare media la examenul de admitere la facultate, notele şi calificativele la examenele şi probele de verificare susţinute în anul universitar, în timp ce judeţul din care provin sau hobby-urile ar putea fi variabile neesenţiale).

De asemenea, variabilele statistice sunt de numeroase tipuri, putând fi clasificate după următoarele criterii:

a) după dimensiunea în care sunt definite, avem:
- *de timp* (care se referă la dimensiunea temporală, de tipul anului înfiinţării unor firme sau al duratei necesare pentru completarea unui formular);
- *de spaţiu* (care se referă la loc, la spaţiu, de tipul localităţii de reşedinţă a unor persoane sau al zonei de amplasare a sediului unor firme);
- *atributive* (care se referă la un atribut, altul decât timpul sau spaţiul, de tipul greutăţii corporale, sexului sau profesiei unor persoane, cifrei de afaceri a unor agenţi economici).

b) după natura variabilelor şi modul lor de exprimare, avem:
- *variabile cantitative* sau *numerice* (exprimate prin numere, adică descriu prin numere rezultatul unei numărători sau măsurători; de exemplu, profitul unor firme, vârsta în ani împliniţi a unor persoane etc.);
- *variabile calitative* sau *nenumerice* (exprimate prin cuvinte, care descriu prin cuvinte un anumit tip calitativ al unităţii de la care s-a înregistrat; de exemplu: liceul absolvit, limba străină cunoscută cel mai bine de nişte persoane etc.)

Scanned with OKEN Scanner

Există situații în care datele sunt măsurate în expresie numerică (cantitativă) și redate în formă calitativă (de exemplu, măsurarea și exprimarea gradului de poluare a aerului: poluarea este măsurată numeric, dar catalogată ca „redusă", „medie" sau „mare"; la fel și măsurarea intensității cutremurelor pe scara Richter).

c) după tipul variației, variabilele numerice pot fi:

- cu *variație continuă*, atunci când pot lua, practic, orice valoare într-un interval din domeniul lor de valori;
- cu *variație discontinuă* (*discrete*), atunci când pot lua doar anumite valori, strict determinate într-un interval din domeniul lor de valori.

În cele mai multe cazuri, deosebirea, diferența între variabilele cu variație continuă și cele cu variație discretă se poate face stabilind dacă datele provin dintr-o numărătoare sau dintr-o măsurătoare (primele sunt discrete, cele din urmă continue). De exemplu, numărul de bolnavi internați într-un spital la 31.12.2005, numărul școlilor generale existente într-un oraș – sunt variabile discrete, deoarece pot lua doar valori întregi, în timp ce greutatea și înălțimea unor persoane, media generală cu care au promovat anul I studenții unei facultăți etc. – sunt variabile continue, întrucât pot lua și valori fracționare. Multe variabile continue pot lua valori aparent discrete (de exemplu, vârsta se exprimă, de regulă, în ani împliniți), dar acest lucru se întâmplă deoarece în practică se limitează, uneori, precizia datelor înregistrate.

În unele cazuri, întâlnim variabile discrete, care pot lua valori foarte mari (de exemplu, populația continentelor la 31.12.2005. Datorită diferenței mari între ordinul de mărime a datelor și pasul cu care se discontinuizează variabila (adică 1, în exemplul nostru), ultimul devine nesemnificativ în raport cu primul, astfel că variabila capătă caracter continuu, chiar dacă ea are, prin însăși natura ei, variație discretă.

d) după numărul de variante pe care-l poate avea o variabilă, întâlnim:

- *variabile alternative*, *binare* sau *dihotomice*, de tipul variabilelor logice („adevărat/fals"), care pot avea doar două variante de răspuns (de exemplu, stagiul militar – satisfăcut/nesatisfăcut, starea civilă – căsătorit/necăsătorit, sexul – masculin/feminin etc.)
- *variabile nealternative*, sunt cele care pot avea cel puțin trei variante de răspuns, de manifestare (de exemplu, salariul – 500 RON, 700 RON, 1000 RON).

Variabilele nealternative se pot transforma în variabile alternative printr-un *proces de dihotomizare* (prin impunerea unei limite, a unui prag – în cazul variabilelor numerice, sau prin regrupări ale variantelor – în cazul variabilelor nenumerice). De exemplu, salariul: mai mic sau egal cu 500 RON și mai mare de 500 RON; profesia: economist și alte profesii).

Scanned with OKEN Scanner

e) după modul de obținere, pot exista:

- *variabile primare* (sunt variabilele care se obțin, de obicei, direct din procesul de culegere a datelor, de la unitățile colectivității statistice);
- *variabile derivate* (sunt variabilele rezultate în urma aplicării unui model de calcul asupra caracteristicilor primare).

f) după relația de cauzalitate, putem avea:

- *variabile independente* – reprezintă variabile ce pot fi utilizate pentru a descrie sau a explica variațiile, diferențele (modificările) survenite în alte variabile, pe care le influențează;
- *variabile dependente* – un rezultat ce prezintă un anumit interes, observat și măsurat pentru a evalua efectele unei variabile independente (de exemplu, dacă pentru 30 de magazine s-au înregistrat suprafața comercială și valoarea vânzărilor, atunci suprafața comercială este variabila independentă, cauzală, iar valoarea vânzărilor reprezintă variabila dependentă, rezultativă).

Relația de cauzalitate între variabila independentă și cea dependentă va fi studiată mai pe larg în capitolul 5.

8. **Varianta**

Nivelul sau valoarea unei variabile, indiferent de natura/tipul acestei variabile, întâlnită la o anumită unitate se numește variantă. De exemplu, un angajat al unei companii are o vechime de 5 ani în muncă. Angajatul reprezintă unitatea statistică, vechimea în muncă este variabila urmărită, iar 5 este varianta acestei variabile.

9. **Frecvența unei variante**

Este numărul de apariții al acelei variante, la toate unitățile unei colectivități. De exemplu, dacă 10 angajați ai unei companii sunt de profesie „economiști", atunci: angajatul este unitatea statistică, profesia este variabila studiată, „economist" reprezintă o variantă a acestei variabile, iar 10 este frecvența acestei variante.

10. **Date statistice**

Reprezintă observațiile rezultate dintr-o cercetare statistică, sau ansamblul valorilor colectate în urma unei cercetări statistice. Conținutul, esența datelor statistice formează ***informația statistică***. În funcție de tipul variabilelor ale căror valori le reprezintă, datele statistice se pot clasifica în: *temporale*, *spațiale* sau *atributive* (dacă se referă la o caracteristică de spațiu, de timp sau atributivă); *cantitative* sau *calitative* (dacă se referă la o variabilă numerică sau nenumerică); datele cantitative (numerice) pot fi: *continue* sau *discrete*. Uneori și datele calitative

Scanned with OKEN Scanner

(sau categoriale) pot fi exprimate numeric, atunci când, de exemplu, li se acordă coduri numerice, ce permit prelucrarea mai facilă cu ajutorul metodelor statistice şi a tehnicii moderne de calcul. Exemplu: codificarea numerică a preferinţelor consumatorilor pentru un produs alimentar, pe o scară de la (1) = deloc gustos, până la (10) = foarte gustos.

În plus, în funcţie de numărul variabilelor la care se referă, datele statistice pot fi:

- *date univariate*, atunci când datele statistice analizate se referă la o singură variabilă sau caracteristică statistică (de exemplu, profitul fiecărui agent economic dintr-o ramură de activitate);
- *date bivariate*, atunci când datele statistice se referă la două variabile; în acest caz, se poate studia separat fiecare caracteristică în parte, sau se poate analiza una dintre variabile, în relaţie cu cealaltă variabilă, punându-se în evidenţă posibila legătură dintre ele (de exemplu, profitul şi forma de proprietate a fiecărui agent economic dintr-o ramură de activitate);
- *date multivariate*, atuci când datele statistice se referă la trei sau mai multe variabile statistice; şi în acest caz se pot analiza separat datele, pentru fiecare variabilă statistică în parte, sau interdependenţa dintre mai multe caracteristici (de exemplu, profitul, forma de proprietate, numărul de ani de activitate, numărul de angajaţi, numărul de oraşe în care au filiale fiecare dintre agenţii economici ce activează într-o ramură). În acest ultim exemplu, întâlnim date cantitative (profitul, numărul de ani de activitate etc.) şi date calitative (forma de proprietate); cele numerice pot fi continue (profitul) sau discrete (numărul de angajaţi, sau numărul oraşelor în care agenţii economici au filiale).

După momentul sau perioada de timp la care se referă datele înregistrate, putem întâlni:

- *date dinamice*, atunci când datele se referă la mai multe momente sau perioade de timp succesive (de exemplu, profitul trimestrial şi numărul mediu trimestrial al angajaţilor unui agent economic din construcţii, înregistrate pentru opt trimestre consecutive);
- *date statice*, atunci când ele se referă la acelaşi moment/perioadă de timp (de exemplu, profitul şi numărul angajaţilor pentru 10 agenţi economici, la sfârşitul semestrului I, 2005).

Scanned with OKEN Scanner

1.5. Scale de măsurare a datelor statistice

Pentru ca datele statistice să poată fi supuse prelucrării prin metode statistice, ele trebuie măsurate, adică exprimate cantitativ, cu ajutorul numerelor. Aşadar, *măsurarea este operaţia prin care fiecărei observaţii, fiecărui aspect cantitativ sau calitativ al unităţilor statistice (obiectelor, evenimentelor etc.) li se atribuie în mod sistematic numere*. Modalitatea prin care se atribuie această valoare numerică, precum şi semnificaţia ei depinde de ***nivelul de măsurare***. Fiecare nivel de măsurare este caracterizat de o anumită precizie, existând 4 nivele principale de măsurare a datelor care se pot ordona după gradul de precizie pe care-l oferă. Fiecărui nivel de măsurare îi corespunde o anumită ***scală de măsurare***, care poate fi identificată în funcţie de operaţiile matematice ce pot fi efectuate cu datele statistice. Toate operaţiile matematice permise pentru o anumită scală sunt valabile şi pentru toate scalele inferioare ei, dar nu şi pentru cele următoare.

Cele patru scale de măsurare, în ordinea crescătoare a preciziei, sunt:

1. ***Scala nominală*** sau ***categorială*** este scala cu cel mai scăzut nivel al preciziei, utilizată pentru a atribui variantelor caracteristicii măsurate „însuşiri", cu scopul de a stabili diferenţe calitative între observaţii; altfel spus, „numerele" atribuite pe această scală au rolul de a identifica, de a reprezenta diferitele clase sau categorii, fără a impune o anumită relaţie de ordine între ele, fără ca aceste „numere" (însuşiri) să aibă vreo semnificaţie cantitativă. Scala nominală are doar rolul de a încadra unităţile statistice în grupe/clase/categorii diferite, după un anumit criteriu, fără a indica dacă o categorie este mai bună sau mai puţin bună decât alta, dacă o categorie este mai largă sau mai restrânsă decât alta.

Exemple: clasificarea persoanelor intervievate în cadrul unui studiu statistic, după statutul civil (căsătorit/necăsătorit), după sex (masculin/feminin), după profesie (inginer/economist/arhitect/avocat etc.). Pentru uşurarea procesului de prelucrare a datelor, aceste variante/ categorii nominative se pot codifica numeric, sub forma: (0 = masculin, 1 = feminin); (1 = inginer, 2 = economist, 3 = arhitect etc.). Aceste numere nu sunt purtătoare ale vreunei informaţii cantitative, ci doar separă subiecţii în clase/categorii distincte din punctul de vedere al statutului civil, al sexului, al profesiei etc. La fel cum numerele de pe tricourile unor jucători de fotbal nu reflectă importanţa jucătorilor sau nivelul la care evoluează în timpul unui meci, ci în principal au rolul de a identifica jucătorii. Cu aceste coduri numerice nu se pot determina medii sau alţi indicatori statistici, nu se pot efectua comparaţii. Putem însă să determinăm frecvenţele de apariţie ale fiecărei categorii (numărul de unităţi care se încadrează în fiecare clasă/categorie).

Exemple: numărul de bărbaţi şi numărul de femei; numărul de persoane căsătorite sau necăsătorite; numărul de ingineri, economişti etc.

Scanned with OKEN Scanner

Singurele relații matematice permise pe scala nominală sunt echivalența și non-echivalența. Astfel, o unitate statistică poate avea caracteristica definitorie pentru o anumită clasă (=) sau nu (≠). Putem spune, de exemplu, „masculin" (≠) „feminin", dar nu și „masculin" > „feminin".

2. ***Scala ordinală*** oferă un plus de precizie în „măsurare", față de scala nominală, deoarece ea nu numai că împarte unitățile colectivității studiate în clase/grupe omogene diferite, din punct de vedere al unei caracteristici, ci permite și stabilirea unei relații de ordine între aceste clase/grupe. În acest caz, valorile numerice atribuite pe această scală pot avea niște numere de ordine (numite și „ranguri"), între care se pot scrie relații în termenii unor „inegalități": $a<b$ sau $a>b$. Astfel, relațiile de ordine întâlnite între categoriile acestei scale pot fi: mai mare sau mai mic, mai rapid sau mai lent, mai mult sau mai puțin inteligent, mai mult sau mai puțin important etc. Aceste expresii numerice sau nenumerice, ce constituie categoriile scalei, conțin mai multă informație cantitativă decât în cazul scalei nominale, dar mai puțină decât informația numerică inclusă în cazul scalelor următoare. Scala ordinală indică poziția unui element într-o serie ordonată, nu și magnitudinea sau mărimea diferenței existente între două poziții ale scalei.

Exemple: popularitatea de care se bucură candidații la o funcție politică (1 = deloc popular, 10 = foarte popular); preferințele consumatorilor față de o marcă de băuturi răcoritoare (1 = nu este deloc bună; 5 = foarte bună); ordinea în care sosesc înotătorii dintr-o cursă acvatică (primul, al doilea, al treilea etc., fără a putea spune cu câte minute și secunde a sosit primul înotător mai repede decât al doilea); calificativele obținute de școlarii claselor I-IV la o materie (insuficient, bine, foarte bine); autopercepția unor persoane privind greutatea lor corporală (sub greutatea normală, cu greutate normală, peste greutatea normală); salariul obținut de angajații unei companii (sub salariul mediu, egal cu salariul mediu, peste salariul mediu); temperatura înregistrată într-o stațiune montană în 10 zile din luna august (sub cea normală, normală, peste cea normală sau foarte scăzută, scăzută, normală, ridicată, foarte ridicată).

Această scală nu ne permite să afirmăm cu câte „unități de măsură" diferă o categorie de categoria învecinată (de exemplu, cu câte grade Celsius diferă temperatura „scăzută" de cea „normală", sau cu câte kilograme este mai grea o persoană din categoria „peste greutatea normală" decât una din categoria „cu greutatea normală").

3. ***Scala de interval (sau cardinală)***

Dacă primele două scale s-au adresat în general variabilelor calitative, scala de interval este prima scală cu adevărat numerică și se aplică variabilelor numerice. Ea aduce în plus față de scala ordinală faptul că permite, pe lângă stabilirea unei relații de ordine între variantele numerice ale acestei scale, și

Scanned with OKEN Scanner

determinarea şi interpretarea diferenţelor dintre acestea. Valorile numerice acordate variantelor caracteristicii pe această scală au semnificaţie cantitativă, unitatea de măsură utilizată este constantă, de aceea este permisă însumarea sau scăderea lor. De exemplu, între anii 1990 şi 1993 există o diferenţă de aceeaşi mărime ca şi între anii 2050 şi 2053.

Pe scala de interval localizarea, fixarea punctului de origine (zero) poate fi făcută arbitrar. Originea nu este, deci, fixă, unică.

De exemplu, pe această scală se măsoară punctajul obţinut la un test de inteligenţă. O persoană, X, care a obţinut la aceste test 60 de puncte a totalizat cu 30 de puncte mai mult decât o altă persoană, Y, care a obţinut doar 30 de puncte, dar nu putem spune că persoana X este de două ori mai inteligentă decât persoana Y; pe această scală nu este, aşadar, permisă, multiplicarea sau divizarea valorilor, ele neavând nici un sens.

Am văzut că o altă trăsătură caracteristică a scalei de interval derivă din posibilitatea alegerii arbitrare a punctului de origine. Rezultă, de aici, că punctul „zero" nu înseamnă, neapărat, şi absenţa caracteristicii, ci este doar o stare, un nivel al acesteia, ca oricare altul. Un exemplu foarte lămuritor este acela al măsurării temperaturii. Să presupunem că avem caracteristica „temperatura medie zilnică într-o localitate în 20 de zile consecutive din luna decembrie". Dacă în una din zile s-a înregistrat o temperatură de 0 grade Celsius, acest lucru nu înseamnă că în acea zi nu a existat temperatură. Extinzând acest raţionament, putem spune că dacă am înregistrat reacţia sau atitudinea unor persoane faţă de un act normativ recent intrat în vigoare, codificarea valorică „zero" a atitudinii nu înseamnă absenţa totală a atitudinii, aşa cum o atitudine codificată cu valoarea 10 nu e de două ori mai puternică decât o atitudine codificată cu valoarea 5. Dacă un elev obţine un punctaj „zero" la un test de verificare a cunoştinţelor, nu înseamnă absenţa totală a oricăror cunoştinţe sau abilităţi de gândire, de raţionament. Un alt exemplu de variabilă măsurabilă pe această scală este timpul calendaristic. Astfel, ora din zi la care adorm 200 de copii ai unor familii se cuantifică cu ajutorul scalei de interval. Ora zero nu înseamnă absenţa orei, ci este doar un nivel, ca oricare altul, al caracteristicii, o manifestare a acesteia.

4. ***Scala de raport*** (sau ***proporţională***) este ultima scală de măsurare, care asigură cel mai înalt grad de precizie şi este, de asemenea, tot o scală numerică. Două valori, măsurate pe această scală se află, indiferent de unitatea de măsură folosită, în acelaşi raport una faţă de alta. Pe această scală sunt permise şi operaţiile de multiplicare şi de divizare. Aşadar, dacă un agent economic are 20 de angajaţi şi un altul are 10 angajaţi, putem afirma că primul are de 2 ori mai mulţi angajaţi decât cel de-al doilea, sau dacă o persoană cântăreşte 90 de kg şi o alta doar 30 de kg, se poate spune că a doua persoană cântăreşte de 3 ori mai puţin decât prima persoană.

Scanned with OKEN Scanner

Diferența între scala de interval și cea de raport este aceea că punctul de origine (zero) este unul fix, rigid, este zero absolut, matematic și reprezintă absența caracteristicii. Nu este neapărat nevoie ca o unitate statistică dintre cele studiate să înregistreze valoarea zero a caracteristicii, este suficient ca această valoare să existe pe scală, fie că ea este atinsă sau nu. De exemplu, vârsta se măsoară cu scala de raport, deși nu există nici o persoană care să aibă 0 ani (adică 0 ani, 0 luni, 0 zile etc.). Timpul după care participanții la un concurs de atletism parcurg distanța de 400 m se măsoară pe scala de raport, deși nu există nici o persoană care să străbată această distanță în 0,00 minute.

În practică este posibil ca o aceeași variabilă să se măsoare cu ajutorul mai multor scale de măsurare. Alegerea scalei adecvate se face, în funcție de variantele acelei caracteristici. Dacă avem caracteristica „salariu", ne gândim imediat la o scală numerică (scala de raport, de exemplu). Într-adevăr, dacă variantele sunt: 750 RON, 1200 RON, 500 RON etc. vom utiliza pentru măsurare scala de raport. Dacă însă variantele sunt: „sub salariul mediu", „egal cu salariul mediu", „peste salariul mediu", atunci vom utiliza o scală calitativă: scala ordinală.

Vom sintetiza cele prezentate în Tabelul 1.1:

Tabelul 1.1.

Denumirea scalei	Caracteristici ale scalei	Exemple de utilizare
NOMINALĂ	Absența relației de ordine	Sex, stare civilă, stagiul militar, profesia, culoarea părului
ORDINALĂ	Relație de ordine	Calificative pentru activitatea școlară, ordinea sosirii alergătorilor dintr-o cursă, preferințele consumatorilor pentru un produs
DE INTERVAL (CARDINALĂ)	Relație de ordine Diferențe (intervale) semnificative Origine arbitrară Valoarea 0 nu înseamnă absență	Temperatura Timpul calendaristic Punctajul obținut la un test de verificare a cunoștințelor
DE RAPORT (PROPORȚIONALĂ)	Relație de ordine Diferențe (intervale) semnificative Origine fixă, valoarea 0 înseamnă absență Operații de multiplicare/ divizare	Vârsta, greutatea corporală, salariul, profitul, cifra de afaceri, numărul de angajați

Scanned with OKEN Scanner

1.6. Procesul de investigație statistică și etapele sale

Economiștii utilizează metode științifice pentru a înțelege comportamentul și mecanismele din mediul economic. În general, o cercetare științifică se bazează pe trei trăsături comune:

(1) încercarea de a răspunde la întrebări, de a soluționa problemele empirice;
(2) utilizarea unor cunoștințe verificabile, sigure, de încredere;
(3) utilizarea empirismului sistematic.

(1) *Rezolvarea unor probleme empirice*

Însuși termenul de „cercetare", de „investigare" descrie o suită de activități prin care lansăm întrebări referitoare la lumea din jurul și din interiorul nostru; primul pas este descrierea, într-o manieră sistematică, a ceea ce studiem, prin procesul de cercetare încercăm să previzionăm și în final să controlăm o anumită situație. Pentru aceasta avem nevoie de date, de informații obiective, absolut necesare pentru a formula o concluzie și a lua o decizie.

Statistica reprezintă un ansamblu de metode de analiză a datelor, dar ea nu trebuie aplicată la întâmplare, ci pe baza unui program. Programul cercetării este un plan sistematic de colectare a datelor în scopul găsirii unor răspunsuri la anumite întrebări, probleme. Alcătuirea planului depinde chiar de problema pe care vrem s-o rezolvăm, plan care, la rândul său, va avea ca rezultat elaborarea unor anumite statistici. Dar cercetarea și analiza corespunzătoare trebuie să țină seama întotdeauna de existența unor condiții de risc, de nesiguranță, răspunsurile găsite vor fi valabile cu o anumită probabilitate. Tocmai datorită acestei incertitudini care însoțește, practic, fiecare latură a vieții, cercetarea sistematică, combinată cu proceduri statistice auxiliare, se dovedește a fi un instrument conceptual și practic atât de puternic. O cercetare realizată după un plan riguros, bine întocmit, în conformitate cu procedurile de analiză statistică va asigura reducerea incertitudinii, a riscului, fără a le putea elimina total. În acest fel, vor putea fi luate decizii mai bine fundamentate.

(2) *Utilizarea unei baze de cunoștințe sigure, verificabile*

Succesul unei cercetări poate fi asigurat, pe lângă întocmirea unui program riguros și prin utilizarea unor definiții operaționale în măsurarea variabilelor supuse analizei. Definiția operațională este mai obiectivă, mai riguroasă și este folosită de cercetător pentru a descrie procesele prin care un obiect sau un eveniment sunt măsurate. Unul din aspectele cele mai importante ale unei definiții operaționale este acela că permite altui cercetător reproducerea, refacerea unui experiment. Deseori, rezultatele la care ajung cercetătorii sunt modificate sub influența unor factori aleatorii, ceea ce face ca și concluziile să poarte amprenta

Scanned with OKEN Scanner

acestor factori. De aceea, este de dorit, uneori, ca studiile să fie refăcute. Cu cât e posibil mai mult acest lucru, cu atât va fi mai mare gradul de încredere al rezultatelor studiului şi convingerea că aceste rezultate sunt reale şi sigure.

Există două modalităţi de refacere a proceselor de cercetare: refacerea directă şi refacerea sistematică.

Prin refacere directă experimentul este repetat în condiţii identice şi are ca scop examinarea siguranţei unui rezultat al studiului. Cu cât repetăm mai des un experiment şi obţinem acelaşi rezultat, cu atât este mai mare încrederea în acest rezultat.

Refacerea sistematică se referă la o repetare generală a unui experiment, dar cu unele modificări substanţiale ale variabilelor independente. Scopul este acela de a examina, de a studia mai detaliat relaţia dintre variabilele independente şi cele dependente.

(3) *Utilizarea empirismului sistematic*

Acest lucru înseamnă că observaţiile pe care le facem trebuie realizate într-o manieră controlată. Astfel, încercăm să formulăm ipoteze alternative şi să determinăm dacă fenomenul pe care îl studiem este determinat de unele condiţii specifice sau de alţi factori.

Produsul final al oricărui proces de cercetare statistică îl reprezintă ansamblul de informaţii rezultate. Aşa cum am văzut, dacă aceste informaţii îmbracă o formă numerică, atunci ele constituie date statistice. Cercetătorul trebuie să sistematizeze, să centralizeze, să interpreteze aceste date, astfel încât, după prelucrarea lor, să obţinem soluţii valide şi sigure la problemele apărute. Aici intervine rolul statisticii, prin analiza datelor şi obţinerea de concluzii pertinente, prin deducerea unor inferenţe pe baza acestor date.

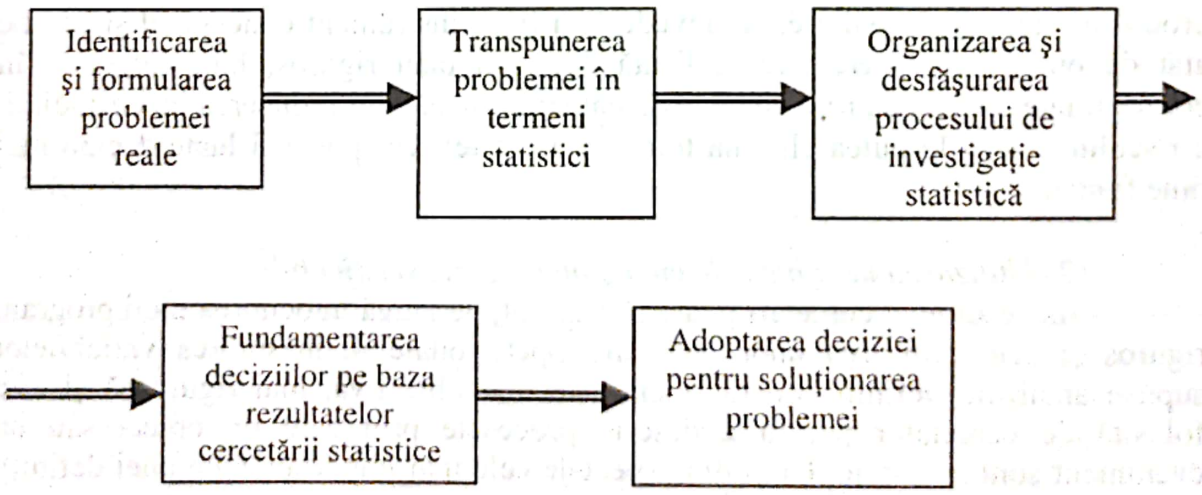


Figura 1.1. Locul cercetării statistice în cadrul procesului decizional

Scanned with OKEN Scanner

Procesul de cercetare statistică presupune parcurgerea următoarelor etape, succesive şi distincte, în funcţie de scopul şi obiectivele urmărite:

- culegerea şi înregistrarea datelor (observarea statistică);
- prelucrarea datelor;
- analiza şi interpretarea rezultatelor,

astfel încât să se obţină un volum suficient de informaţii, cu un grad de exactitate acceptabil, utilizându-se, totodată, un minim de resurse materiale, financiare şi umane.

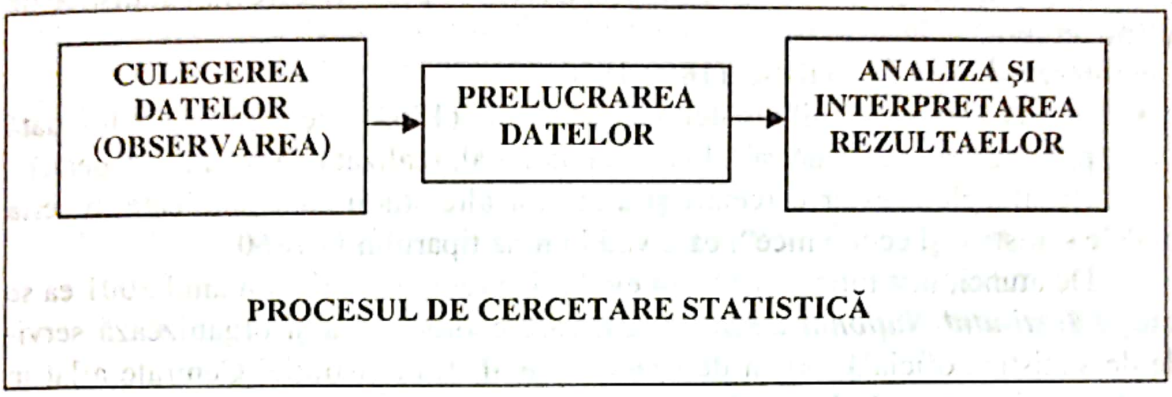


Figura 1.2. Etapele procesului de cercetare statistică

În etapa de ***observare statistică*** are loc culegerea şi înregistrarea datelor de la unităţile colectivităţii, referitoare la toate variabilele studiate, pe baza unui program riguros şi sistematic.

În etapa de ***prelucrare*** datele sunt sistematizate, centralizate şi sunt calculaţi indicatori statistici ce caracterizează toate laturile fenomenului urmărit.

În etapa de ***analiză şi interpretare*** are loc compararea rezultatelor obţinute, verificarea ipotezelor, formularea concluziilor şi fundamentarea calculelor de prognoză, cu alte cuvinte se construieşte fundamentul ştiinţific pe baza căruia se vor adopta deciziile manageriale care să ducă la soluţionarea problemei reale.

1.7. Organizarea statisticii oficiale în România şi pe plan internaţional

„Credem cu seriozitate că statistica este, pentru cei care o posedă, o preţioasă avere naţională. Statistica este singurul tărâm sigur pe care un Guvern prevăzător trebuie să-şi întemeieze raţiunea reformelor sale." (Dionisie Pop Marţian).

„Naşterea" statisticii oficiale în România a avut loc prin semnarea, de către Domnitorul Alexandru Ioan Cuza, a Ordonanţei de înfiinţare a Oficiului Central de

Scanned with OKEN Scanner

Statistică Administrativă. Contribuții marcante și-au adus atunci doi economiști importanți ai veacului al XIX-lea: Dionisie Pop Marțian și Ion Ionescu de la Brad.

Prima instituție a Statisticii oficiale și-a circumscris încă de la început activitatea principiilor de adevăr, de cunoaștere și evaluare corectă a realităților țării, oferind, prin informații și date statistice, fundamentul necesar Guvernului, agenților economici pentru elaborarea și aplicarea unor programe realiste de guvernare, de dezvoltare economico-socială.

Încă din primii ani de existență, cu insuficiente resurse materiale, financiare și umane specializate – instituția a realizat primele cercetări statistice de mare anvergură:

- recensământul populației (1859-1860);
- recensământul stabilimentelor industriale (1863), ce constituie totodată prima cercetare de acest fel la nivel național, realizată de o țară europeană.

Rezultatele acestor cercetări și a multor alte studii sunt publicate în seria „Analele statistice și economice", care văd lumina tiparului în 1860.

De atunci, instituția a cunoscut evoluții și reorganizări. Din anul 2001 ea se numește ***Institutul Național de Statistică***, care coordonează și organizează serviciile de statistică oficială, organ de specialitate al Administrației Centrale aflat în subordinea Guvernului. Instituția desfășoară activități de culegere, prelucrare, analiză, difuzare și constituirea seriilor de date statistice oficiale cu caracter demografic, social, economic, financiar, juridic, necesare fundamentării și evaluării politicilor economice și sociale, deciziilor guvernamentale și ale agenților economici, informării opiniei publice, elaborării cercetărilor științifice, prognozelor și strategiilor de dezvoltare, transmiterii de date statistice către organismele internaționale, potrivit obligațiilor asumate de statul român, precum și altor categorii de utilizatori interni și externi.

În subordinea Institutului Național de Statistică funcționează direcțiile regionale de statistică, organizate la nivelul județelor – centre ale regiunilor de dezvoltare și direcțiile județene de statistică, organizate la nivelul celorlalte județe, ca instituții publice cu personalitate juridică.

Cadrul legislativ în care se desfășoară activitatea de statistică oficială în România este asigurat de OG nr. 9/1992 privind organizarea statisticii oficiale, modificată și completată prin OG nr. 67/2003, OG nr. 71/2004, de HG nr. 68/2004 privind organizarea și funcționarea Institutului Național de Statistică, modificată prin HG nr. 175/2005.

Serviciile de statistică oficială sunt organizate și desfășurate prin respectarea principiilor de autonomie, confidențialitate, transparență, importanță, proporționalitate, deontologie statistică și a raportului cost/eficiență. La acestea se adaugă și „Principiile fundamentale ale Statisticii Oficiale" adoptate la cea de-a 42-a Sesiune a ONU/CEE din aprilie 1992, între care menționăm:

Scanned with OKEN Scanner

- statisticile oficiale trebuie să îndeplinească condiția utilității practice, fiind elaborate şi puse la dispoziția Guvernului, a agenților economici şi a publicului larg pe baze imparțiale de către organismele de statistică oficială publică, facilitând accesul cetățenilor la informațiile publice;
- pentru a menține încrederea în statisticile oficiale, organismele acesteia acționează şi funcționează în conformitate cu principiile strict profesionale, ştiințifice şi ale eticii profesionale, pe baza unor metode şi proceduri de colectare, prelucrare, stocare şi prezentare a datelor statistice;
- pentru a asigura o interpretare corectă a datelor, organismele abilitate ale statisticii oficiale trebuie să prezinte informații privind sursele, metodele şi procedurile statistice utilizate, în conformitate cu standardele ştiințifice;
- organismele abilitate ale statisticii oficiale sunt îndreptățite să ia poziție asupra erorilor de interpretare, asupra utilizării deficitare sau eronate a informațiilor şi datelor statistice;
- datele statistice pot fi obținute, deopotrivă, din cercetări statistice sau înregistrări administrative, fiind la latitudinea instituțiilor abilitate să aleagă sursa, astfel încât datele obținute să întrunească condițiile de calitate, operativitate şi eficiență a costurilor;
- datele individuale colectate de către organismele statisticii oficiale în vederea prelucrării, indiferent că se referă la persoane fizice sau juridice, sunt strict confidențiale şi vor fi folosite exclusiv în scopuri statistice;
- legile, regulile şi măsurile pe baza cărora funcționează sistemul statistic trebuie făcute publice;
- cooperarea bilaterală şi multilaterală în ceea ce priveşte activitatea statistică, cu alte organisme abilitate din alte țări contribuie la îmbunătățirea sistemului statisticilor oficiale din toate țările.

Institutul Național de Statistică poate obține datele şi informațiile necesare pentru studierea fenomenelor şi proceselor economice, sociale, demografice şi de mediu fie prin raportările oficiale ale agenților economici, fie prin cercetări proprii organizate, precum: recensăminte, microrecensăminte, monografii statistice, cercetări statistice selective şi prin alte anchete statistice. De asemenea, serviciile de statistică oficială au acces la datele primare din evidențele administrative ale instituțiilor administrației publice centrale şi locale, inclusiv la datele individuale de identificare, cu respectarea principiului confidențialității statistice. Aceste date se vor utiliza numai în scopuri statistice. Datele stocate în bănci de date vor fi prelucrate, cu ajutorul tehnicilor moderne, utilizându-se sisteme de indicatori, nomenclatoare, clasificări, metodologii, tehnici de prelucrare stabilite în mod imparțial şi independent chiar de către Institutul Național de Statistică, ținându-se seama şi de recomandările organismelor internaționale din domeniu pentru asigurarea comparabilității pe plan internațional a rezultatelor obținute. Rezultatele

Scanned with OKEN Scanner

cercetărilor sunt popularizate printr-o serie de publicaţii pe suport de hârtie sau electronic, în general cu caracter periodic, dar şi ocazional (anuare, buletine statistice informative, studii statistice publicate în Revista Română de Statistică – revista proprie a Institutului), conferinţe, simpozioane, mese rotunde sau alte evenimente la care se poate satisface nevoia de informare a utilizatorilor de date statistice şi a populaţiei, în condiţiile specifice ale protecţiei şi confidenţialităţii datelor şi informaţiilor statistice.

Pentru asigurarea caracterului obiectiv, transparent şi ştiinţific al metodologiilor, indicatorilor, nomenclatoarelor şi clasificărilor utilizate în activitatea de statistică s-a creat ***Consiliul de Coordonare a Activităţii de Statistică***, organ consultativ cu atribuţii în activitatea de analiză şi avizarea strategiei de dezvoltare a sistemului statistic naţional, a rapoartelor de activitate ale Institutului Naţional de Statistică şi a Programului anual de cercetări statistice.

Organismele statisticii oficiale din ţara noastră cooperează în mod permanent cu instituţiile similare ale altor ţări, precum şi cu organismele oficiale pe plan multinaţional care au rolul de a asigura armonizarea, compatibilitatea datelor obţinute.

În cadrul ***Organizaţiei Naţiunilor Unite*** (ONU) există un ***Departament de Statistică***, cu atribuţii în:

- colectarea, procesarea şi diseminarea informaţiei statistice;
- standardizarea metodelor statistice, a clasificărilor şi definiţiilor;
- elaborarea şi participarea la programe de cooperare;
- coordonarea programelor şi activităţilor de statistică internaţională.

Pentru realizarea acestor scopuri, Departamentul de Statistică al ONU:

- formează un centru de date privind comerţul internaţional, conturile naţionale, energie, industrie, mediu, transport şi statistici sociale şi demografice, colectate din surse naţionale şi internaţionale;
- elaborează şi promovează standarde internaţionale ale metodelor, clasificărilor şi definiţiilor utilizate de organismele naţionale de statistică;
- acordă asistenţă statelor membre, la cererea acestora, pentru îmbunătăţirea serviciilor lor statistice, atât prin recomandări, cât şi prin training;
- coordonează programele statistice internaţionale încredinţate acestui Departament de către Comisia de Statistică şi de Comitetul de Coordonare a Activităţii Statistice;
- asigură informaţiile şi suportul secretarial pentru Comisia de Statistică a ONU şi pentru grupul de lucru al acesteia;
- ajută la dezvoltarea şi diseminarea statisticilor mediului, a indicatorilor dezvoltării durabile, promovează tehnici moderne de cercetare statistică şi monitorizare, ca instrument al creşterii economice şi dezvoltării.

Scanned with OKEN Scanner

Departamentul de Statistică al ONU editează o serie de publicaţii (între care Anuarul statistic), cărţi şi rapoarte statistice. De asemenea, multe din bazele de date ale Departamentului sunt disponibile pe suport electronic.

În cadrul Organizaţiei Naţiunilor Unite există o ***Comisie de Statistică***, formată din 24 de ţări membre ale ONU, alese de ***Consiliul Socio-Economic al ONU*** pe o perioadă de 4 ani, pe baza unei distribuţii geografice echitabile. Comisia monitorizează activitatea ONU în domeniul statisticii.

Uniunea Europeană (UE) are şi ea un ***Departament de Statistică***, cu rolul de a asigura servicii informaţionale statistice de înaltă calitate (Eurostat). ***Eurostat*** cooperează şi cu ţările care nu sunt membre UE, în misiunea de a coordona optimizarea sistemelor statistice din ţările în curs de dezvoltare, candidate la aderarea la Uniunea Europeană. Au fost stabilite programe speciale cu ţările din Europa Centrală şi Estică (PHARE) şi cu Comunitatea Statelor Independente ale fostei Uniuni Sovietice (TACIS). Eurostat cooperează îndeaproape şi cu ţările mediteraneene şi cu cele africane.

Eurostat a înfiinţat **Comitetul pentru Programe Statistice**, al cărui principal obiectiv este acela de a asigura statistici comparabile la nivelul Uniunii Europene. Acest Comitet include atât Eurostat-ul, cât şi oficiile statistice, ministere, agenţii şi bănci centrale care desfăşoară activitate de statistică oficială în statele membre, în Islanda, Norvegia şi Liechtenstein, într-o structură de reţea în care rolul Eurostat-ului este acela de a conduce activitatea de armonizare a statisticilor oficiale, în strânsă colaborare cu autorităţile statistice naţionale. De asemenea, tot el colaborează şi cu alte organizaţii internaţionale, precum OECD, ONU, Fondul Monetar Internaţional şi Banca Mondială. Principalele sale activităţi vizează:

- dezvoltarea unor clasificări, metodologii şi definiţii comune pentru toate statele membre;
- implementarea unor cercetări selective comune, bazate pe metodologii comune, armonizate;
- colectarea, analiza şi diseminarea datelor statistice pentru Uniunea Europeană, inclusiv asigurarea comparabilităţii între diferite ţări şi regiuni.

Cadrul administrativ al statisticii în unele ţări ale lumii

Belgia

În 1939 a fost creat un serviciu centralizat de statistică, prin Decret Regal, împreună cu unele activităţi statistice subsecvente. În 1985 a luat fiinţă Institutul Naţional de Statistică, cu rol coordonator al sistemului naţional statistic. El este departament al Ministerului Afacerilor Economice, metodele de cercetare statistică

Scanned with OKEN Scanner

sunt elaborate de Institutul Național de Statistică, în cooperare cu Înaltul Consiliu al Statisticii.

Canada

Are un sistem al statisticii oficiale centralizat foarte extins, deservit de un Birou Național de Statistică. Bugetul lui este autorizat de Parlament, la recomandarea Trezoreriei. Programele statistice sunt elaborate în colaborare cu Consiliul Național Statistic.

Franța

Sistemul francez al statisticii oficiale este relativ descentralizat, cu servicii statistice incluse în mai multe ministere și agenții de statistică, publice sau semipublice. Agenția Centrală, numită Institutul Național de Studii Economico-Sociale, creat printr-un act legislativ al Parlamentului (1946) are funcția coordonatoare. Consiliul Național de Informație Statistică, condus de ministrul afacerilor economice asigură cooperarea între furnizorii (producătorii) și utilizatorii informației statistice și elaborează recomandări asupra modului de desfășurare ale activității statistice.

Germania

Sistemul statisticii oficiale din Germania este preponderent federalizat, activitatea fiind coordonată de Biroul Federal Statistic și de Birourile Statistice Regionale (ale landurilor). Există și alte agenții care realizează unele studii în diverse domenii. Președintele Biroului Federal Statistic este conducătorul unui comitet consultativ, care acordă asistență Biroului în problemele fundamentale. Biroul are independență în ceea ce privește metodologia științifică utilizată.

Italia

În 1988 au fost stabilite în mod legal criteriile și principiile care stau la baza serviciilor statistice italiene. Serviciile de statistică oficială au un caracter pronunțat descentralizat, însă întregul sistem național de statistică este coordonat de Institutul Național de Statistică, al cărui buget este, sub aspect legal și administrativ, autonom. Comisia pentru Protecția Informației Statistice monitorizează asigurarea imparțialității, obiectivității în activitatea statistică, a calității metodelor statistice și a tehnicilor de prelucrare a datelor și asigură cooperarea cu Uniunea Europeană.

Olanda

Sistemul statisticii oficiale din Olanda este complet centralizat la nivel național, aproape toate informațiile sunt prelucrate și diseminate de către Biroul

Scanned with OKEN Scanner

Central de Statistică, înființat prin decret regal în 1899. El se află sub coordonarea Ministerului Afacerilor Economice, dar în ceea ce privește activitatea statistică, este complet independent. Biroul, în colaborare cu Comisia Centrală de Statistică (ai cărei membri sunt numiți prin decret regal) elaborează un program al activității statistice; implementarea programului constituie în exclusivitate responsabilitatea Biroului Central.

Spania

În 1989 a fost adoptată o lege a statisticii, care reglementează Statistica oficială din această țară. Motorul sistemului îl constituie Institutul Național pentru Statistică, care coordonează activitatea de statistică desfășurată și în departamentele de profil din cadrul ministerelor și autorităților regionale. Institutul Național pentru Statistică funcționează ca agenție autonomă pe lângă Ministerul Economiei și Finanțelor. Acest institut elaborează, în cooperare cu părțile descentralizate ale sistemului, proiectul Programului Național Statistic. Înaltul Consiliu Statistic, condus de ministru oferă asistență în elaborarea programului.

Suedia

Sistemul statisticii oficiale în Suedia are caracter descentralizat; fiecare agenție guvernamentală cu atribuții în ceea ce privește statistica în domeniul său de interes. Rolul coordonator îi aparține, însă, Agenției Centrale de Statistică. Activitatea de statistică oficială este reglementată de un act legislativ al statisticii generale din 1992. Consiliul Științific are un rol consultativ pe lângă Agenția Centrală.

Statele Unite ale Americii

SUA are un sistem statistic descentralizat. Câteva departamente ale Guvernului Federal au incluse agenții de statistică, iar alte 60 de agenții federale se ocupă cu publicarea statisticilor. Oficiul de Management și Buget din cadrul Biroului Executiv al Președintelui este responsabil cu politica în domeniul statisticii și cu coordonarea activității de statistică oficială. O serie de comitete congresionale supervizează activitățile Oficiului, iar Comitetul Național de Statistică al Academiei Naționale de Știință are rol consultativ în domeniu.

Scanned with OKEN Scanner

CAPITOLUL 2
CULEGEREA ŞI PRELUCRAREA PRIMARĂ A DATELOR STATISTICE

2.1. Observarea statistică

2.1.1. Importanţa şi conţinutul observării statistice

Proiectarea unei cercetări statistice este o operaţiune de mare amploare, antrenând un volum considerabil de resurse materiale, financiare şi umane şi un sistem informaţional bine articulat, cu funcţionalitate permanentă. Organizarea sistemului informaţional depinde întotdeauna de rolul şi funcţiile pe care le are statistica în procesul de conducere şi de fundamentare a deciziilor, atât la nivel macro, cât şi în structurile componente ale sistemului. Totodată, buna funcţionare, în condiţii de eficienţă maximă a sistemului informaţional depinde şi de logistica, de tehnica modernă de calcul de care se dispune. Sistemul este conceput într-o structură ierarhică, pe baza elementelor lui obţinându-se indicatori agregaţi de la bază către vârf şi indicatori dezagregaţi, de la vârful sistemului către bază.

În vederea cunoaşterii şi conducerii fenomenelor şi proceselor economico-sociale cu ajutorul metodelor şi tehnicilor statistice, este necesară întâi caracterizarea, descrierea cât mai fidelă a situaţiei existente. Acest lucru se realizează prin strângerea, colectarea datelor care să exprime starea de fapt, adică prin observarea statistică.

Orice proces de investigaţie statistică începe cu observarea statistică, modul în care este ea organizată şi realizată influenţează, în continuare, celelalte etape ale procesului de cercetare statistică, deoarece obţinerea unor date autentice, reale, exacte determină calitatea etapelor de prelucrare, analiză şi intepretare a rezultatelor.

Datele necesare se pot obţine din diferite surse:

a) din investigaţii statistice special organizate, atunci când datele nu există iniţial sub forma necesară – acestea sunt ***datele primare***;
b) din sistemul informaţional existent (de exemplu, din evidenţele curente ale agenţilor economici, bănci şi baze de date pe suport de hârtie sau electronic ale unor instituţii ale administraţiei centrale şi locale etc.) şi se numesc ***date secundare***.

Scanned with OKEN Scanner

Observarea statistică este prima etapă a investigației statistice, care constă în culegerea și înregistrarea în mod sistematic și unitar a datelor statistice, de la unitățile colectivității, referitoare la toate caracteristicile incluse în programul cercetării.

Observarea statistică nu se desfășoară la întâmplare, deoarece culegerea datelor implică anumite costuri, dacă vrem să obținem date autentice. Ea se desfășoară pe baza unui plan (program), elaborat în funcție de scopul cercetării statistice. Observarea trebuie fie corespunzătoare cantitativ și calitativ.

- îndeplinirea *condiției de cantitate* (de volum) înseamnă obținerea în timpul prestabilit a întregului volum de date necesar pentru atingerea tuturor obiectivelor investigației statistice;
- îndeplinirea *condiției de calitate* presupune asigurarea autenticității, a caracterului real al datelor, astfel încât rezultatele obținute prin prelucrarea acestor date să fie cât mai exacte și să ducă la adoptarea unor decizii cât mai pertinente.

În cazul în care pentru obținerea datelor statistice este necesară organizarea unei investigații statistice, atunci ea trebuie făcută în conformitate cu un anumit plan, pentru a ne asigura că obținem, cu un minim de resurse materiale și financiare, rezultate cât mai exacte și mai pertinente.

2.1.2. *Programul observării* (în accepțiune mai largă) cuprinde următoarele elemente:

- scopul observării;
- colectivitatea observată;
- unitatea de observare;
- timpul și locul observării;
- lista caracteristicilor urmărite;
- formularele de înregistrare a datelor;
- măsuri organizatorice.

Programul observării, într-o accepțiune restrânsă, cuprinde doar o listă a tuturor caracteristicilor ce vor fi înregistrate, a indicatorilor necesari etc.

A. ***Scopul observării*** este subordonat scopului general al cercetării statistice și influențează în continuare toate celelalte elemente din programul observării. Stabilirea obiectivelor ce trebuie atinse prin investigarea organizată este punctul de pornire al elaborării programului.

B. ***Delimitarea populației*** studiate presupune identificarea, stabilirea tuturor unităților statistice ce vor fi incluse în cercetare și de la care se vor înregistra valorile caracteristicilor. Acest lucru se realizează cu ajutorul nomenclatoarelor, al clasificărilor existente, sau se apelează la studiile anterioare efectuate în domeniul de interes (dacă există). Gradul de întindere a colectivității de observare este în funcție de metoda de observare ce va fi utilizată: dacă este vorba de o metodă de

Scanned with OKEN Scanner

observare totală, atunci obiectul observării va fi format din toate unităţile colectivităţii, iar dacă va fi aplicată o metodă de observare parţială, vor fi înregistrate date doar de la o parte a unităţilor colectivităţii şi doar ele vor forma obiectul observării. În ambele cazuri însă, pentru identificarea corectă şi completă a colectivităţii ce va fi supusă observării este necesară determinarea coordonatelor de spaţiu, timp şi de organizare ale colectivităţii de interes.

C. ***Unitatea de observare*** este unitatea la care se referă caracteristicile înregistrate, *simplă* (dacă este vorba de un salariat, un obiect etc.) sau *complexă* (dacă se are în vedere o echipă sau o familie). Uneori, unitatea de observare nu este totodată şi unitate raportoare. De exemplu, în cazul unei cercetări efectuate în cadrul unui agent economic, unitatea raportoare va fi agentul economic, iar unitatea de observare poate fi angajatul, un departament, o echipă sau un produs.

D. Stabilirea ***timpului observării*** impune precizarea a două aspecte:

- timpul la care se referă datele înregistrate (care poate fi un moment, numit „moment critic", sau o perioadă, un interval de timp; în primul caz fenomenul este surprins static, iar în cel de-al doilea caz – dinamic).
- timpul când se efectuează înregistrarea datelor – care este, de regulă, un interval de timp, bine precizat; este de dorit ca acesta să fie cât mai scurt, astfel încât înregistrarea datelor să se realizeze cu o cât mai mare operativitate.

De exemplu, la Recensământul populaţiei şi locuinţelor din martie 2002, momentul critic a fost ora 0 a zilei de 18 martie, iar perioada în care au fost înregistrate datele a fost 18-27 martie. Unitatea de observare a fost gospodăria (unitate complexă).

Locul observării este, de regulă, locul unde se produce, unde are loc fenomenul urmărit şi studiat.

E. ***Lista caracteristicilor urmărite***, care formează, aşa cum am văzut, programul (planul) observării în sens restrâns, presupune stabilirea tuturor variabilelor care trebuie înregistrate, astfel încât să se asigure studierea fenomenului de interes sub toate aspectele prevăzute în scopul cercetării. În acelaşi timp, însă, trebuie să se evite redundanţa informaţiilor.

Caracteristicile pot apărea sub diferite forme:

- sub forma unor indicatori – în raportările statistice pe care le completează agenţii economici;
- sub forma răspunsurilor la întrebările dintr-un chestionar – în cazul investigaţiilor statistice special organizate.

Pentru a obţine toate datele necesare cercetării şi a îndeplini cu succes scopul observării, proiectarea chestionarelor trebuie făcută astfel încât să se asigure o structură logică şi o formulare bine gândită a întrebărilor. Prin structura sa, chestionarul trebuie să realizeze o trecere firească, o succesiune logică de la o

Scanned with OKEN Scanner

întrebare la alta, de la un subiect de discuţie la altul, deoarece dacă s-ar trece brusc şi în mod contrar logicii fireşti de la un subiect la altul, intervievatul (respondentul) ar fi dezorientat şi ar fi influenţat în acordarea răspunsului. Deseori, se foloseşte ca tehnică generală trecerea de la întrebările generale către cele specifice sau referitoare la un anumit subiect.

Organizaţia Gallup este de părere că prin formularea unei întrebări într-un chestionar, trebuie să se atingă cinci obiective:

1. Să se asigure că respondentul cunoaşte tema asupra căreia se realizează cercetarea şi că nu este în afara ei.

Exemplul 2.1. „Ştiaţi de planurile privind construcţia unui centru de afaceri în zona X?"

2. Să obţină o atitudine generală din partea respondentului privitoare la problema în cauză.

Exemplul 2.2. „Ce credeţi despre necesitatea construirii acestui centru de afaceri?" (răspuns pe o scală ordinală, cu variantele: „Sunt întru totul de acord", „Sunt de acord", „Nu sunt sigur", „Nu sunt de acord", „Sunt cu totul împotrivă"), sau „Credeţi că acest centru de afaceri este necesar în zonă?" (Da/Nu).

3. Să obţină răspunsuri privind un aspect particular, specific al problemei.

Exemplul 2.3. „Credeţi că noul centru de afaceri va afecta mediul înconjurător din zonă?" (Da/Nu).

4. Să obţină motivaţii ale răspunsurilor date de cei chestionaţi, ale punctelor lor de vedere asupra problemei.

Exemplul 2.4. „Dacă sunteţi împotriva construirii centrului de afaceri, atunci motivul dumneavoastră ar fi:

a) există şi aşa o mare densitate de clădiri în zonă;
b) construcţia ar dăuna peisajului existent în zonă;
c) proiectul ar însemna demolarea unei construcţii de interes naţional sau a unui parc de joacă pentru copii";
d) alt motiv şi anume .....

5. Să afle în ce măsură şi-ar susţine persoana intervievată punctul de vedere.

Exemplul 2.5. „Ce acţiune credeţi că aţi întreprinde pentru a sensibiliza autorităţile asupra punctelor dumneavoastră de vedere?"

a) să scrieţi unui consilier local;
b) să scrieţi unui parlamentar;
c) să semnaţi o petiţie;
d) să luaţi cuvântul la o dezbatere publică pe tema respectivă;
e) să participaţi la o demonstraţie;
f) altă acţiune şi anume .....

Într-un chestionar putem întâlni întrebări deschise (în care respondentul poate da orice răspuns, de tipul: „Ce profesie aveţi?") sau întrebări închise (în care

Scanned with OKEN Scanner

respondentului îi sunt indicate o serie de posibile răspunsuri, din care poate alege unul sau mai multe, de genul: „Câte persoane are familia dumneavoastră?" cu variantele: 1, 2, 3, 4, 5 sau mai multe).

Întrebările mai pot fi, de asemenea, factuale (de tipul: „Ce număr purtați la pantofi?") sau de opinie (de tipul: „Ce părere aveți despre intenția Guvernului de a majora TVA-ul?", cu variantele de răspuns: „Nu sunt de acord", „Sunt de acord", „Nu știu").

F. ***Formularele de înregistrare*** reprezintă suportul material pe care vor fi înregistrate datele culese; formularele pot fi de două tipuri: formulare tip fișă sau formulare tip listă.

- în formularul tip *fișă* se înregistrează datele referitoare la o singură unitate statistică (de aceea fișa este un formular individual); fișa se utilizează atunci când unitățile statistice de la care se înregistrează sunt mai răspândite, mai dispersate în profil teritorial, sau atunci când programul observării este mai bogat (cuprinde mai multe caracteristici ce vor fi înregistrate de la fiecare unitate statistică);
- în formularul tip *listă* se vor înregistra date pentru mai multe unități de observare (formular colectiv); acest tip de formular se folosește cu precădere atunci când unitățile statistice sunt mai concentrate teritorial (când se regăsesc multe unități statistice într-o arie teritorială restrânsă) sau când programul observării este mai sumar (cuprinde mai puține caracteristici ce trebuie înregistrate de la fiecare unitate statistică).

De exemplu, la recensământul populației și locuințelor din martie 2002 s-a folosit fișa pentru a înregistra datele referitoare la o gospodărie și lista pentru a consemna datele privind militarii în termen dintr-o unitate militară sau persoanele care trăiesc într-un cămin de bătrâni, sau studenții dintr-un cămin studențesc.

Pentru asigurarea clarității întrebărilor dintr-un formular, acesta poate fi însoțit de instrucțiuni (tipărite direct pe formular sau într-o anexă a acestuia), astfel încât respondenții să completeze formularul cu răspunsuri cât mai reale și mai fidele realității. Pentru ca datele oferite ca răspuns să fie prelucrate în mod automat, se utilizează formulare tipizate, iar răspunsurile vor fi codificate.

G. ***Măsuri organizatorice***. Se referă la totalitatea măsurilor ce trebuie luate pentru a se asigura desfășurarea în bune condiții, conform programului stabilit, a observării statistice. Cu ajutorul lor, se pot preîntâmpina și rezolva diversele obstacole, dificultăți ce s-ar putea ivi în realizarea observării. Aceste măsuri pot cuprinde: întocmirea unor liste ale unităților ce vor fi supuse observării, folosirea hărților și planurilor cu localizarea unităților statistice, sectorizarea teritoriului, recrutarea și instruirea personalului participant la colectarea datelor, proiectarea, elaborarea și tipărirea formularelor de înregistrare, popularizarea cercetării statistice (prin mass-media, afișe etc.).

Scanned with OKEN Scanner

De exemplu, la organizarea unui recensământ, acțiunea trebuie popularizată pentru a atrage buna credință a populației, astfel încât aceasta să furnizeze date corecte și reale; uneori, pentru identificarea și soluționarea tuturor problemelor ce ar putea apărea, se organizează un recensământ de probă.

2.1.3. Metode de observare statistică

Există o gamă variată de metode de observare, utilizate împreună sau separat, pentru a obține o informație statistică cât mai reală, completă și neredundantă. Metodele de observare se clasifică după următoarele criterii:

1. După *frecvența* înregistrării în timp a datelor despre fenomenele și procesele economico-sociale, putem avea:
   a. **observări curente** – atunci când fenomenele și procesele sunt urmărite în mod continuu, iar datele ce le caracterizează sunt înregistrate permanent (de exemplu, fenomenele demografice: natalitatea, mortalitatea, nupțialitatea, divorțialitatea care se regăsesc în „statistica stării civile"; rezultatele obținute de agenții economici etc.);
   b. **observări periodice** – atunci când datele referitoare la fenomene sau procese economico-sociale se înregistrează cu o frecvență aproximativ regulată, la intervale egale de timp (de exemplu, recensăminte ale populației, recensăminte agricole etc.);
   c. **observări ocazionale** – atunci când datele referitoare la fenomenele sau procesele economico-sociale se înregistrează discontinuu, ocazional, cu un prilej sau scop special, în mod unic, motiv pentru care se mai numesc „special organizate" (de exemplu, o anchetă de opinie privind atitudinea cetățenilor referitoare la un nou act normativ recent intrat în vigoare; o anchetă de opinie privind preferințele participanților la o expoziție de modă etc.).
2. După *gradul de cuprindere* a populației de observare delimitate, întâlnim:
   a. **observări totale** – atunci când vor fi supuse observării toate unitățile populației statistice (de exemplu, recensământul populației și locuințelor);
   b. **observări parțiale** – atunci când vor fi supuse observării doar o parte a unităților populației statistice (de exemplu, anchete, sondaje statistice).
3. După *modul de obținere a datelor*, avem:
   c. **observări primare** (directe) – în care datele se obțin prin culegerea, înregistrarea lor direct de la unitățile statistice (de exemplu,

Scanned with OKEN Scanner

recensământul populației, o anchetă de opinie);

d. **observări secundare** (indirecte) – în care se obțin datele secundare, care se preiau din documentele existente (de exemplu, datele preluate din evidențele contabile).

Principalele *metode* utilizate *de observare statistică* sunt:

a) ***Recensământul statistic*** – o metodă de observare totală, periodică, prin care un fenomen este surprins în mod static, înregistrându-se valorile caracteristicilor din programul observării referitoare la un moment de referință numit „moment critic", de la toate unitățile populației statistice.

Recensămintele sunt întâlnite din timpuri străvechi, încă din Antichitate, când romanii, egiptenii efectuau recensăminte ale populației; din sfera populației, recensământul s-a extins și în sfera locuințelor, animalelor, unităților din industrie, agricultură, comerț („recensământ economic"). În domeniul populației recensământul are o periodicitate de 10 ani, iar în domeniul economic de 5 ani. Recensământul este un tip de înregistrare foarte laborios, necesitând un volum mare de cheltuieli, personal numeros, măsuri organizatorice detaliate, precum și o prelucrare complexă a informațiilor înregistrate.

Recensământul populației este una din principalele surse de date în statistica demografică și asigură informații privind numărul și structura populației de pe teritoriul unei țări la un moment dat. El este inițiat de către stat, fiind reglementat printr-un act normativ, însă responsabilitatea pentru organizarea și punerea sa în practică revine Institutului Național de Statistică, care asigură și prelucrarea datelor înregistrate. Recensământul populației se efectuează în baza respectării principiilor universalității, simultaneității și comparabilității.

Respectarea principiului universalității presupune înregistrarea, cu ocazia Recensământului populației, în mod obligatoriu, a tuturor persoanelor aflate în jurisdicția statului respectiv. Astfel, la Recensământul populației și locuințelor din martie 2002 au fost înregistrați toți cetățenii români cu domiciliul în țară, indiferent dacă la momentul de referință se aflau pe teritoriul țării sau erau plecați temporar în străinătate, precum și persoanele de altă cetățenie sau fără cetățenie, care aveau domiciliul sau reședința temporară în România.

Conform principiului simultaneității, informațiile culese reflectă situația existentă la un același moment de timp pentru toate persoanele recenzate (momentul critic), deși înregistrarea datelor durează un anumit interval de timp. Momentul critic este ales, de obicei, iarna, atunci când populația observată se află într-un moment de stabilitate, uniformitate, nu este supusă unor fluctuații, unor mișcări accidentale (în acest sens, se evită vacanțele școlare sau concediile de odihnă).

Scanned with OKEN Scanner

De exemplu, la Recensământul populației și locuințelor din 18 martie 2002, momentul critic a fost ora zero a zilei de 18 martie, iar înregistrarea datelor s-a realizat în perioada 18-27 martie.

Pe baza Recensământului populației obținem indicatori de volum și structură statici, aferenți momentului critic. O cercetare în domeniul populației are însă, ca obiectiv, și surprinderea, determinarea modificărilor de volum și a mutațiilor de structură de la un Recensământ la altul, impunându-se a fi respectat principiul comparabilității datelor atât în plan temporal, cât și în plan teritorial (armonizarea sferei de cuprindere a indicatorilor, a definițiilor, nomenclatoarelor, clasificărilor utilizate la mai multe recensăminte succesive, precum și între țări, ținându-se seama și de recomandările organismelor internaționale de specialitate).

Perioada de pregătire a recensământului este de mare amploare și răspundere, fiind necesară o investigare a modalităților de pregătire a recensămintelor din perioadele trecute, atât în România, cât și în alte țări, o mediatizare puternică a acțiunii, uneori chiar și organizarea unui recensământ de probă.

b) ***Sondajul statistic*** (***selecția statistică***) este o metodă de observare parțială, special organizată, una din cele mai des utilizate metode de observare statistice moderne, deoarece realizează o economie de resurse materiale, financiare și umane, înlocuind o observare totală, de mare extindere, mai dificil de organizat și efectuat, cu cheltuieli considerabile. De asemenea, cu ajutorul sondajului, se asigură culegerea și prelucrarea datelor într-un timp mai scurt, în mod operativ.

În cazul cercetării statistice prin sondaj, se selectează un eșantion, format dintr-o parte a unităților populației totale. Eșantionul poate fi selectat aleatoriu sau non-aleatoriu. *Selectarea aleatorie* a unităților în eșantion presupune că fiecare unitate statistică are o șansă determinabilă (calculabilă), egală și nenulă de a fi inclusă în eșantion, că nu există deci nici o unitate care să nu poată fi supusă procedeului de selecție și că nici o unitate nu poate fi aleasă preferențial. Eșantioanele alese aleatoriu sunt reprezentative pentru întreaga populație, adică întrunesc aceleași trăsături esențiale ca și populația totală.

Selectarea non-aleatorie (dirijată) se referă la orice metodă de alegere, pe baza unui criteriu preferențial, a unităților statistice în eșantion. Odată format eșantionul, se înregistrează nivelurile tuturor caracteristicilor incluse în programul observării, de la toate unitățile acestuia, se prelucrează și se obțin indicatori statistici la nivelul eșantionului (etapa descriptivă), pentru ca apoi să se extindă, cu o anumită probabilitate, rezultatele de la nivelul eșantionului pentru întreaga colectivitate (etapa inferențială). Diferența dintre nivelul estimat la indicatorului în urma sondajului și nivelul real (*eroarea de sondaj*) va fi cu atât mai mică, cu cât eșantionul este mai reprezentativ, iar reprezentativitatea, la rândul ei, va fi asigurată dacă se respectă principiile selecției aleatorii în formarea eșantionului.

Scanned with OKEN Scanner

c) ***Ancheta statistică*** este tot o metodă de observare parţială, mai mult cu caracter ocazional, care diferă de sondajul statistic prin neobligativitatea îndeplinirii condiţiei de reprezentativitate a eşantionului. În plus, în cazul unei anchete, se admite şi o rată mai mare a non-răspunsurilor la formularele înmânate direct, telefonic, prin e-mail sau prin poştă, întrucât completarea chestionarelor este benevolă. Ancheta statistică va conduce, din aceste motive, la erori mai mari decât sondajul statistic. În cazul acestuia din urmă, neobţinerea datelor solicitate de la unităţile incluse în eşantion (persoane, de regulă), din diferite motive, precum: nu cunosc limba oficială, şi-au schimbat domiciliul, nu sunt acasă atunci când sunt căutaţi, refuză să coopereze etc. – pot duce la afectarea reprezentativităţii eşantionului. O categorie deosebit de întâlnită a anchetelor statistice o formează ***anchetele de opinie***, ce au drept scop studierea preferinţelor, atitudinii oamenilor faţă de diferite probleme (anchete în sfera sociologiei, marketingului, politicii, culturii etc.).

Vom prezenta, în continuare, una din cele mai importante sondaje statistice proiectate, organizate şi efectuate de Institutul Naţional de Statistică: ancheta asupra forţei de muncă în gospodării (AMIGO).

Ancheta asupra forţei de muncă în gospodării (AMIGO)

Importanţă şi obiective

Piesă de bază într-un dispozitiv statistic al forţei de muncă adecvat economiei de piaţă, Ancheta asupra forţei de muncă în gospodării (AMIGO) îşi propune, pe de o parte, evaluarea populaţiei active – ocupate sau în şomaj – şi a populaţiei inactive, iar pe de altă parte, determinarea structurilor acestor categorii, precum şi a tendinţelor lor de evoluţie. AMIGO a devenit, astfel, o metodă modernă pentru cercetarea statistică a pieţei forţei de muncă, răspunzând nevoilor de informare în domeniu.

Între două recensăminte ale populaţiei, care au avut loc la intervale mari de timp, această anchetă are drept scop obţinerea unor rezultate intercensitare privind forţa de muncă, dar şi a unor statistici originale asupra ocupării, asupra duratei muncii, a mobilităţii pieţei forţei de muncă din punct de vedere social şi economic, asupra nivelului de instruire şi de formare profesională. Ea permite, de asemenea, o detaliere a situaţiei şomerilor şi evidenţierea diferenţei dintre numărul şomerilor înregistraţi la oficiile de forţă de muncă şi şomaj din reţeaua Ministerului Muncii şi cel determinat potrivit recomandărilor internaţionale (Biroul Internaţional al Muncii).

Bucurându-se de o tradiţie veche în ţările cu economie de piaţă, acest tip de anchetă, având drept univers gospodăriile populaţiei, este unicul instrument prin care se poate măsura complexitatea şi interacţiunea fenomenelor de pe piaţa forţei

Scanned with OKEN Scanner

de muncă. Anchetele legate de forţa de muncă din întreprinderi nu analizează decât un segment al populaţiei ocupate (salariaţii), în timp ce alte categorii ale populaţiei ocupate, precum lucrătorii pe cont propriu, patronii, cei care îşi ajută familia într-o activitate economică sau membrii cooperativelor rămân în afara sferei de cercetare statistică. De asemenea, ancheta în gospodării este şi o cale de obţinere a unor informaţii suplimentare asupra şomajului care, deşi înregistrat prin canale administrative, nu poate fi surprins la adevărata lui dimensiune, pe de o parte pentru că nu toţi şomerii se înscriu ca atare la Oficiile de forţă de muncă şi şomaj, pe de altă parte, pentru că dintre toţi cei înscrişi, unii mai prestează activităţi ocazionale. Alte aspecte importante pentru caracterizarea pieţei forţei de muncă, pe care însă sursele statistice existente nu le pot evalua la adevărata lor dimensiune sunt: ocuparea incompletă (subocuparea) forţei de muncă, situaţia pensionarilor, a altor categorii de persoane inactive, a persoanelor descurajate sau a celor care au pe lângă activitatea principală şi o activitate secundară.

Prin flexibilitatea sa, acest tip de anchetă permite comensurarea simultană a tuturor segmentelor de populaţie şi, în acelaşi timp, a întregii lor problematici, oferind o imagine de ansamblu asupra diverselor fenomene şi procese sociale şi demografice, precum şi a corelaţiilor existente în cadrul acestora.

În anii 1994 şi 1995, ancheta s-a desfăşurat o dată pe an. Începând cu anul 1996, ancheta AMIGO furnizează informaţii trimestriale, ceea ce va asigura completarea sistemului informaţional statistic actual cu date conjuncturale asupra şomajului şi populaţiei ocupate. Se vor putea evidenţia şi măsura fenomenele cu caracter sezonier de pe piaţa forţei de muncă. Concepţia metodologică şi organizatorică a anchetei a fost realizată cu asistenţa tehnică a specialiştilor francezi de la Institutul Naţional de Statistică şi Studii Economice (Franţa) în cadrul Programului PHARE în domeniul statisticii. Organizarea practică şi conducerea anchetei în teren s-a făcut de către Comisia Naţională pentru Statistică, în prezent Institutul Naţional de Statistică, prin direcţiile sale judeţene de statistică.

Ancheta furnizează elementele necesare studiului privind evoluţia mărimii şi structurii ofertei de forţă de muncă şi a multiplelor fenomene care se manifestă pe piaţa forţei de muncă, prin compararea atât cu trimestrul anterior, cât şi cu trimestrul corespunzător din anul precedent.

Datorită informaţiilor furnizate, ancheta AMIGO stă la baza elaborării unor programe adecvate de protecţie socială.

Sfera de cuprindere

În sfera de cuprindere a anchetei intră persoanele de cetăţenie română, cele fără cetăţenie sau de cetăţenie străină, cu domiciliul permanent în centrele de cercetare selectate.

Scanned with OKEN Scanner

În eşantion au fost incluse gospodăriile din locuinţele permanente (inclusiv unităţi locuite din necesitate), repartizate în centre de cercetare din toate judeţele ţării, cuprinzând proporţional gospodării din mediul urban şi rural.

Fac obiectul înregistrării toţi membrii gospodăriilor, inclusiv persoanele plecate din localitate pentru o perioadă îndelungată, de peste 6 luni, dacă acestea păstrează legături familiale cu gospodăria din care fac parte, precum militarii în termen, elevii şi studenţii plecaţi la studii, persoanele plecate la lucru, deţinuţii şi arestaţii; persoanele spitalizate sau aflate temporar în sanatoriu pentru tratament sau recuperare. Nu intră în sfera de cercetare persoanele care locuiesc permanent în unităţile de locuit în comun (cămine de bătrâni, de handicapaţi, cămine muncitoreşti, sanatorii etc.).

Metoda de înregistrare

Metoda de înregistrare a datelor în chestionare este interviul, luat de către anchetatori special instruiţi, persoanelor de 15 ani şi peste, la domiciliul gospodăriilor din locuinţele cercetate.

Informaţiile obţinute ca răspuns la întrebările directe adresate membrilor gospodăriilor se înscriu în chestionarele anchetei: chestionarul individual şi chestionarul locuinţei.

Perioada de înregistrare şi perioada de referinţă

Interviurile au fost repartizate uniform de-a lungul trimestrului, realizându-se o cercetare continuă.

Perioada de referinţă (pentru marea majoritate a întrebărilor) a fost săptămâna (de luni până duminică, inclusiv) dinaintea înregistrării. Pentru anumite întrebări există şi alte perioade de referinţă: ultimele 4 săptămâni, ultimul an, următoarele 15 zile, ultimele 3 luni din anul curent sau din perioada corespunzătoare a anului precedent etc.

Eşantionul anchetei

Ancheta se realizează selectiv, pe un eşantion aleatoriu, reprezentativ pentru întreaga ţară, utilizând un plan de sondaj în două trepte. Eşantionul este reprezentativ atât din punctul de vedere al structurii gospodăriilor, cât şi al distribuţiei populaţiei pe medii, sexe şi grupe de vârstă.

În prima treaptă a planului de sondaj s-a construit Eşantionul *Multifuncţional de Zone Teritoriale* (eşantionul „master" EMZOT), constituit din 501 de centre de cercetare – unităţi primare de eşantionare – repartizate în toate judeţele ţării şi sectoarele Municipiului Bucureşti) şi care se actualizează periodic. În a doua treaptă de formare a eşantionului se aleg locuinţele (18.036 locuinţe cu toate

Scanned with OKEN Scanner

gospodăriile care le compun), luându-se în considerare obținerea unor estimații ale principalelor caracteristici cercetate, care să fie afectate de erori în limita a 3% și garantate cu o probabilitate de 97%.

Eșantionul trimestrial este ales după o schemă rotațională prin care o locuință este cercetată două trimestre consecutiv, următoarele două trimestre ies temporar din cercetare, apoi revine în cercetare în următoarele două trimestre, după care iese definitiv din cercetare, deci o locuință este gestionată 6 trimestre. Conform schemei aplicate („schema rotațională 2-2-2"), în fiecare trimestru se păstrează în eșantion 50% din locuințele trimestrului anterior, 25% din locuințele cercetate cu 2 trimestre în urmă, iar restul de 25% sunt locuințe absolut noi. Astfel, se păstrează și o acoperire de 50% a eșantioanelor de la un trimestru al unui an la același trimestru al anului precedent. Utilizarea acestei scheme prezintă următoarele avantaje: asigură micșorarea ratei de non-răspuns și a erorilor ce pot rezulta din anchetările repetate, evită menținerea în cercetare a aceloraşi grupuri de locuințe, pentru perioade lungi de timp, asigură echilibrarea cerințelor de calitate pentru estimatorii caracteristicilor cercetate la un moment dat și în timp, reduce discontinuitățile în seriile de timp.

Extinderea rezultatelor

În fiecare trimestru începând din anul 2002, extinderea rezultatelor obținute din anchetă, la nivelul întregii țări, s-a realizat pe baza ponderilor atribuite persoanelor care fac parte din eșantion, care au răspuns la interviu. La calcularea acestor ponderi s-a avut în vedere rata de non-răspuns trimestrială și distribuția populației legale pe sexe, grupe de vârstă, medii de rezidență și regiuni: la 1 ianuarie 2002 pentru trimestrele I și II, respectiv la 1 iulie 2002 pentru trimestrele III și IV. Ratele de răspuns au variat între 94,4% în trimestrul I și 95,0% în trimestrul IV.

Datele anuale extinse pentru anul 2002 s-au obținut ca medie aritmetică a datelor trimestriale extinse. *Datele pentru anul 2002 au fost estimate pe baza rezultatelor Recensământului Populației și Locuințelor din martie 2002.*

Extinderea rezultatelor obținute din anchetă la nivelul întregii țări s-a realizat pe baza coeficienților atribuiți persoanelor aparținând gospodăriilor din eșantion, care au răspuns la interviu.

d) ***Panelul*** este o metodă de observare parțială, realizată pe baza unui eșantion fix, stabil, format din aceleași unități care răspund la aceleași întrebări, la mai multe momente succesive, diferite de timp (tehnica longitudinală). În acest fel cercetarea se concentrează pe identificarea, pe determinarea modificărilor survenite de-a lungul timpului asupra fenomenului studiat. Această metodă prezintă, însă, și unele dezavantaje, precum:

Scanned with OKEN Scanner

- respondenții se implică, treptat, în cercetare și, știindu-se vizați, sub observație, și-ar putea schimba comportamentul și răspunsurile lor ar putea fi influențate;
- odată cu trecerea timpului, o parte dintre respondenți părăsesc panelul (mortalitate), iar cei rămași sunt mai puțin reprezentativi în populația totală.

De exemplu, se pot organiza cercetări statistice prin panel în următoarele cazuri: pentru determinarea impactului unor campanii publicitare susținute asupra unei populații, pentru monitorizarea preferințelor electorale ale unui grup de cetățeni etc.

e) ***Observarea părții principale*** (***metoda masivului principal de date***) este, de asemenea, o metodă de observare parțială, special organizată (cu caracter ocazional). Metoda presupune formarea unui eșantion nereprezentativ, prin selectarea, dintr-o populație împărțită pe grupe de importanță diferită, a unităților mai semnificative (din grupele cu pondere mai însemnată). Înregistrând și prelucrând datele de la unitățile incluse în masivul principal de date se pot estima suficient de corect valorile unor indicatori pentru întreaga colectivitate statistică.

De exemplu, pentru estimarea nivelului veniturilor salariale ale angajaților unei mari companii (dintre care o parte sunt angajați permanent și o parte sunt colaboratori), vor fi incluși în studiu numai cei din prima categorie, deoarece ei dețin ponderea cea mai mare din totalul personalului.

f) ***Monografia*** este o metodă de observare parțială, special organizată, care presupune caracterizarea detaliată, sub multiple aspecte, ale unei singure unități statistice complexe (se elaborează monografia unui unei întreprinderi, a unui județ, a unei localități etc.). Monografia include, pe lângă culegerea și înregistrarea datelor și aspecte legate de prelucrarea lor și de analiză și intepretare a rezultatelor, toate cele trei etape ale investigației statistice putând fi realizate de o echipă complexă de specialiști.

2.1.4. Conceptul de „eroare statistică". Modalități de control al datelor statistice

Deoarece statistica operează cu seturi mari de date, este posibilă apariția, după fiecare etapă a procesului de investigație statistică, a unor erori. Putem întâlni, așadar, erori de observare (de înregistrare), erori de prelucrare, erori de analiză și de interpretare.

În accepțiune largă, *eroarea statistică reprezintă diferența dintre nivelul real al unui indicator și nivelul obținut în urma procesului de investigație statistică.*

Eroarea statistică se poate determina în expresie absolută sau relativă.

a) ***Eroarea în expresie absolută*** se determină în sensul definit mai sus, conform relației:

Scanned with OKEN Scanner

$$e_x = x - x_0, \qquad (2.1.)$$

unde e_x este mărimea absolută a erorii statistice;

x este nivelul indicatorului determinat în urma investigației statistice;

x_0 este nivelul real al indicatorului.

De cele mai multe ori, însă, nivelul real al indicatorului nu este cunoscut, deci nu se poate calcula eroarea absolută. Ea se poate interpreta, însă, cu ajutorul teoriei probabilităților.

Eroarea în expresie absolută are unitate de măsură concretă, aceeași cu cea a indicatorului/caracteristicii pentru care se calculează, de aceea nu este permisă compararea mai multor erori absolute, decât dacă ele sunt determinate pentru un același indicator/caracteristică. Pentru a elimina dependența de unitatea de măsură, se calculează eroarea în expresie relativă.

b) ***Eroarea în expresie relativă*** se determină ca raport între eroarea absolută și nivelul real al indicatorului și se măsoară în procente:

$$e_x^{\%} = \frac{e_x}{x_0} = \frac{x - x_0}{x_0} \cdot 100 \qquad (2.2.)$$

Principalele tipuri de erori statistice sunt:

a) ***Erori de observare*** (sau ***de înregistrare***)

Apar în etapa de observare statistică și sunt definite ca diferențele care apar între valorile reale ale caracteristicilor urmărite (studiate) și cele înregistrate (culese); sunt determinate fie de obiectul supus observării, de personalul care efectuează înregistrarea datelor, de mijloacele sau metoda de observare sau de alți factori. Erorile de observare sunt de două tipuri:

a_1) *erori întâmplătoare* – sunt, în general, erorile produse din neatenție, sau din alte cauze accidentale, independent de voința anchetatorului sau a respondentului. De regulă, acest tip de erori se produc în ambele sensuri față de nivelul real al indicatorului (altfel spus, valoarea înregistrată a indicatorului/caracteristicii poate fi mai mare sau mai mică decât valoarea reală a acestuia). Conform experienței practice, deoarece în cercetările statistice se culeg date de la un număr foarte mare de unități statistice (creându-se câmp de acțiune Legii numerelor mari), a rezultat că la nivelul ansamblului, erorile aleatoare se compensează reciproc, neafectând calitatea datelor.

a_2) *erori sistematice* – au caracter premeditat, se realizează, de regulă, cu bună știință, uneori chiar cu voința anchetatorului sau a respondentului. Acest tip de erori se produc, în general, într-un singur sens față de valoarea reală a fenomenului studiat și de aceea nu se mai compensează reciproc la nivel total, afectând calitatea datelor. Dintre cauzele care ar putea duce la producerea erorilor sistematice, menționăm neînțelegerea sau interpretarea incorectă a chestionarelor

Scanned with OKEN Scanner

trebuie completate, atât din partea anchetatorului, cât şi a respondentului, comoditatea sau reaua voinţă a persoanei care efectuează culegerea şi înregistrarea datelor, neînţelegerea, nerespectarea scopului cercetării, delimitarea greşită a populaţiei de observare, definirea incorectă a caracteristicilor urmărite, formularea neclară a întrebărilor din chestionare, utilizând cuvinte complicate sau ambigue, mutarea unor răspunsuri dintr-o perioadă în alta etc. De exemplu, influenţarea de către anchetator a răspunsurilor persoanelor intervievate, furnizarea de către respondenţi a unor răspunsuri inexacte, neconforme cu realitatea, deoarece răspunsurile reale l-ar pune într-o lumină nefavorabilă etc.

Erorile de observare pot fi eliminate sau reduse prin control statistic.

b) ***Erori de reprezentativitate***

Se întâlnesc în cazul utilizării sondajului statistic, atunci când selectarea unităţilor în eşantion nu se face în mod aleatoriu, ci după un criteriu subiectiv, preferenţial şi, la rândul lor, se împart în sistematice şi aleatorii. Ele nu pot fi total eliminate, ci doar diminuate şi controlate.

c) ***Erori de prelucrare***

Apar în etapa de prelucrare a datelor statistice, din cauza utilizării greşite a unor metodolgii – *erori metodologice*, (de exemplu, reprezentarea datelor statistice cu ajutorul unor tipuri neadecvate de grafice, gruparea sau clasificarea greşită, utilizarea unor formule de calcul nepotrivite, care vor duce la obţinerea unor rezultate eronate etc.), utilizarea greşită a unor modele statistice (*erori de modelare* – apar în procesul de elaborare, de alegerea greşită a modelelor).

Pentru obţinerea unor rezultate corecte ale cercetării, este bine ca după fiecare etapă a investigaţiei statistice, datele statistice să fie supuse unor operaţii de control, în vederea identificării şi eliminării (sau măcar a minimizării) unor eventuale erori. În etapa de observare, se poate aplica asupra datelor culese un control cantitativ şi calitativ.

Controlul cantitativ al datelor presupune verificarea completitudinii datelor culese şi înregistrate, prin:

- verificarea primirii tuturor chestionarelor la centrul de prelucrare;
- verificarea completării tuturor rubricilor din chestionare.

Controlul cantitativ al datelor se realizează atât în cazul observărilor statistice totale, cât şi în cel al observărilor statistice parţiale (în cazul utilizării sondajului statistic, de exemplu, existenţa unui volum mare de date incomplete sau inexistente – non-răspunsuri – poate afecta reprezentativitatea eşantionului).

Pe de altă parte, trebuie eliminate situaţiile de redundanţă a datelor, întrucât acestea ar duce la cheltuieli suplimentare nejustificate.

Scanned with OKEN Scanner

Controlul calitativ al datelor se referă la asigurarea unui nivel calitativ corespunzător al datelor culese. Poate îmbrăca două forme:

- *controlul aritmetic*, care presupune aplicarea prin sondaj, a unor relații de calcul între diferiții indicatori înscriși în chestionare, de tipul egalităților de balanță sau al altor chei de verificare;
- *controlul logic* presupune verificarea corectitudinii datelor înregistrate pe baza unor legături, identități logice între mai multe valori ale aceleiași caracteristici sau între valorile unor caracteristici diferite. Pentru aceasta se pot utiliza diferite deducții logice, rezultatele unor studii asemănătoare sau experiența acumulată de personalul specializat (de exemplu, între valoarea mediei și valorile individuale din care a fost calculată, între vârstă, nivelul de educație și profesie, sau între vârstă și starea civilă etc.).

Corectarea erorilor, după ce acestea au fost depistate, se poate face atât pe cale *manuală*, cât și în mod *automat*, cu ajutorul tehnicii de calcul.

La *corectarea manuală*, o valoare depistată ca eronată este înlocuită cu una plauzibilă, rezonabilă, ce va fi aleasă de către o persoană.

La *corectarea automată*, valoarea eronată este înlocuită în mod automat cu o valoare plauzibilă, compatibilă cu celelalte valori, găsită de calculator pe baza unui algoritm (de exemplu, algoritmul de „căutare a celui mai apropiat vecin"). La fel, se poate proceda și în cazul datelor lipsă (a non-răspusurilor), în locul acestora fiind, de asemenea, generate valori posibile, care se înscriu în anumite limite și sunt necontradictorii cu restul datelor.

2.2. Prelucrarea primară a datelor statistice

Totalitatea operațiilor de sistematizare a datelor statistice, de prezentare și reprezentare a acestora, de centralizare și de aplicare a unor modele simple de determinare a unor indicatori sintetici, alcătuiesc *prelucrarea primară*. Desigur, în prelungirea acestui concept, se pot realiza, în continuare, **analize mai complexe, a dependențelor existente între variabilele statistice, a tendințelor de evoluție în timp a fenomenelor, a identificării și cuantificării influenței factorilor asupra unui fenomen complex, toate acestea, împreună cu operațiile de prelucrare primară, formează *prelucrarea în sens larg*.**

Așa cum am arătat anterior, statistica descriptivă implică organizarea într-o manieră ordonată a datelor statistice, sintetizarea, centralizarea lor, prezentarea și reprezentarea lor astfel încât să permită evidențierea într-un timp operativ, a caracteristicilor esențiale ale setului de date. Metodele statisticii descriptive reunesc, deopotrivă, tehnici de reprezentare grafică, dar și modalități descriptive,

Scanned with OKEN Scanner

numerice, de măsurare (prin care se obțin indicatori sintetici, precum media sau dispersia).

2.2.1. Sistematizarea datelor statistice: clasificarea și gruparea datelor statistice

Am văzut că statistica operează cu volume mari de date. Dacă aceste date sunt prezentate într-o formă aleatorie neregulată, este dificil, investigând vizual setul de date, să-l caracterizăm prin trăsăturile sale esențiale, prin valorile extreme, tendința centrală sau gradul de dispersare. De aceea, putem întâi supune setul de date unor *operații de prezentare sistematică, de organizare, de ordonare a acestor date după unul sau mai multe criterii*, într-un cuvânt, de ***sistematizare***.

Această operație face trecerea de la observarea statistică (în urma căreia s-au obținut datele necesare realizării obiectivelor investigației statistice, verificate sub aspectul volumului și calității) și prelucrarea propriu-zisă.

Să presupunem că au fost culese și înregistrate valorile a m caracteristici $X_1, X_2, ..., X_m$, de la r unități statistice. Datele se pot prezenta ca în Tabelul 2.1.:

Prezentarea ordonată a datelor statistice după unități și variabile statistice

Tabelul 2.1.

Unitatea statistică	Caracteristica urmărită					
	X_1	X_2	...	X_j	...	X_m
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1m}
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2m}
...	...	...	...	...	...	...
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{im}
...	...	...	...	...	...	...
r	x_{r1}	x_{r2}	...	x_{rj}	...	x_{rm}
Total	$\sum_{i=1}^{r} x_{i1}$	$\sum_{i=1}^{r} x_{i2}$	...	$\sum_{i=1}^{r} x_{ij}$	...	$\sum_{i=1}^{r} x_{im}$

unde:

i = unitatea statistică;

X_j = variabila urmărită;

x_{ij} = nivelul variabilei X_j înregistrată la unitatea statistică i.

O linie din Tabelul 2.1., de forma (x_{i1} , x_{i2} ,..., x_{ij} ,..., x_{im}) reprezintă valorile tuturor caracteristicilor întâlnite la unitatea statistică i. O coloană din tabel, de forma (x_{1j} , x_{2j} , ...,x_{ij} ,..., x_{rj}) reprezintă valorile unei singure caracteristici (X_j), înregistrate la toate unitățile statistice care posedă acea caracteristică.

Scanned with OKEN Scanner

Desigur că în prezentarea matriceală din Tabelul 2.1. se poate analiza, separat, fiecare coloană a matricii, adică fiecare variabilă statistică în parte, cu variantele sale înregistrate la toate unitățile statistice corespunzătoare, sau se pot analiza, corelat, două sau mai multe coloane (adică două sau mai multe caracteristici), pentru a se identifica posibilitatea existenței unei legături între ele, forma de manifestare a legăturii, direcția sau sensul ei, intensitatea cu care se produce.

Dacă valorile caracteristicii urmărite sunt numerice, cantitative și însumabile, atunci putem calcula și valorile centralizate, agregate, ale fiecărei variabile. De exemplu, dacă ele sunt exprimate în unități fizice de același fel (cantități dintr-un același produs), atunci valorile x_{ij}, $i=\overline{1,r}$ sunt direct însumabile.

Dacă valorile x_{ij}, $i=\overline{1,r}$ ale variabilei X_j sunt exprimate în unități fizice de natură diferită (cantități din diferite produse), atunci, pentru a le face însumabile și a calcula nivelul centralizat, total, al caracteristicii, ele trebuie transformate în valori însumabile cu ajutorul unor coeficienți de echivalență (de exemplu, prețul unei unități din produsul respectiv); astfel, vom obține expresii valorice care pot fi însumate.

Este posibil, însă, ca nivelurile x_{ij}, $i=\overline{1,r}$ ale variabilei X_j să nu poată fi deloc însumate și atunci nu se va determina nivelul total al caracteristicii.

Este clar că, pe măsură ce avem mai multe unități statistice de la care înregistrăm mai multe variabile, cu atât prezentarea tuturor datelor într-o formă matriceală ca cea din Tabelul 2.1. devine mai dificilă și cu atât va fi mai greu să ne facem o viziune de ansamblu asupra întregului set de date. De aceea este indicat să ordonăm datele, să le organizăm, adică să le supunem operațiilor de grupare/clasificare.

***Gruparea/clasificarea* datelor statistice presupune împărțirea unităților populației statistice observate în grupe sau clase distincte omogene, după unul sau mai multe criterii.**

Atunci când criteriul (caracteristica) după care se realizează această operație este unul numeric, cantitativ, ea se numește ***grupare***, iar când operația se realizează după un criteriu (caracteristic) calitativ, nenumeric, ea se numește ***clasificare***.

La realizarea unei grupări/clasificări, trebuie, pe cât posibil, să îndeplinim următoarele condiții:

a) ***omogenitate*** (în sensul că unitățile statistice care au aceeași valoare sau valori apropiate, asemănătoare, ale caracteristicii după care se efectuează sistematizarea vor fi incluse în aceeași clasă; în felul acesta, se dorește ca variația valorilor caracteristicii incluse în aceeași grupă/clasă să fie cât mai mică);

Scanned with OKEN Scanner

b) ***unicitate*** (în sensul că o unitate statistică trebuie inclusă într-o singură clasă sau grupă, ea nu se poate regăsi simultan în două sau mai multe clase/grupe);

c) ***completitudine*** (în sensul că toate unitățile statistice să fie incluse în grupe/clase, să nu fie exclusă vreo unitate din operația de sistematizare).

Sunt cazuri în care nu este posibilă îndeplinirea simultană a tuturor acestor condiții (de exemplu, dacă sunt unități la care s-au înregistrat valori extreme, aberante, ale caracteristicii după care se face sistematizarea, este de dorit, uneori, să se evidențieze separat aceste cazuri, să se scoată în afara grupării aceste unități și să se sistematizeze restul unităților, la care s-au înregistrat valori mai apropiate ale variabilei).

În funcție de *numărul de criterii/variabile* după care efectuăm sistematizarea, putem avea:

- ***grupări/clasificări simple***, atunci când sistematizarea se realizează după valorile unei singure caracteristici numerice sau nenumerice;
- ***grupări/clasificări combinate***, atunci când sistematizarea se realizează după valorile a două sau mai multe variabile;

În funcție de numărul variantelor caracteristicii după care se realizează gruparea/clasificarea, putem avea:

- grupare/clasificare pe variante, dacă variabila după care se efectuează sistematizarea are puține variante numerice sau nenumerice (de exemplu, clasificarea persoanelor după sex: masculin/feminin; clasificarea persoanelor după statutul civil: căsătorit/necăsătorit; gruparea familiilor după numărul de membri: 1/2/3/4/peste 4 etc.)
- grupare/clasificare pe grupe de variante (în cazul unui criteriu calitativ de sistematizare) sau pe intervale de variație (în cazul unui criteriu cantitativ, numeric de sistematizare), atunci când s-au înregistrat mai multe variante ale caracteristicii de ordonare (de exemplu, clasificarea studenților după regiunea de dezvoltare în care își au reședința presupune comasarea județelor de reședință pe regiuni de dezvoltare: regiunea de nord-est, de sud-est, de sud, de sud-vest, de vest, de nord-vest, centrală, București; gruparea angajaților unei companii după nivelul veniturilor salariale: sub 500 RON, 500-600 RON, 600-700 RON, 700-800 RON, 800-900 RON, 900-1000 RON, peste 1000 RON).

Scanned with OKEN Scanner

Atunci când gruparea se efectuează după o caracteristică numerică pe intervale de variaţie, acestea, la rândul lor, pot fi: egale sau neegale, continue sau discontinue.

În funcţie de natura/conţinutul variabilei după care se efectuează sistematizarea, putem avea:

- grupări/clasificări după o ***caracteristică de timp*** (numite şi *grupări/clasificări cronologice*), cum ar fi: gruparea unor firme după anul de înfiinţare, gruparea turiştilor după perioada sejurului;
- grupări/clasificări după o ***caracteristică de spaţiu*** (numite şi *grupări/clasificări teritoriale*), cum ar fi: gruparea magazinelor după sectorul din Bucureşti în care se găsesc;
- grupări/clasificări după o ***caracteristică atributivă***, cum ar fi: gruparea unor posturi de televiziune după veniturile din reclamă şi publicitate).

A. Clasificări statistice

***Clasificarea* presupune repartizarea unităţilor unei colectivităţi pe clase sau categorii, după o variabilă calitativă.** Se construiesc un număr de clase egal cu numărul categoriilor existente, iar numărul unităţilor statistice incluse în fiecare clasă reprezintă *frecvenţa* acelei clase (volumul ei).

Există clasificări care au caracter oficial şi altele cu caracter neoficial.

Institutul Naţional de Statistică, în calitatea sa de coordonator al sistemului de clasificări de interes naţional, a creat un server al nomenclatoarelor de interes naţional (SENIN). Această acţiune se înscrie în strategia de construire a Societăţii Informaţionale din România.

SENIN pune la dispoziţia utilizatorilor clasificările de interes general, alături de principiile metodologice aplicate la construirea clasificărilor, descrierea structurii clasificărilor, prezentarea generală a clasificărilor/nomenclatoarelor, legătura cu alte clasificări/nomenclatoare naţionale şi internaţionale, reguli de încadrare a activităţilor, produselor etc.

Dintre clasificările/nomenclatoarele cu caracter oficial din România amintim: CAEN (Clasificarea Activităţilor din Economia Naţională), PRODROM (nomenclator de produse şi servicii industriale), CPSA (clasificarea produselor şi serviciilor asociate activităţilor din economia naţională), SIRUTA (registrul unităţilor administrativ-teritoriale ale României), SCN (Sistemul Conturilor Naţionale) etc.

CAEN este emis de Institutul Naţional de Statistică, având la bază Hotărâri ale Guvernului. Ultima versiune a fost implementată în anul 2002, sistemul fiind revizuit în funcţie de modificările apărute şi legiferate la nivel european şi este de la început construit în acord cu clasificarea europeană, adaptată la specificul economiei româneşti. Actuala versiune CAEN din România reprezintă o clasificare

Scanned with OKEN Scanner

detaliată a ISIC (Clasificarea Internațională Standard a activităților economice), elaborată de Comisia de Statistică a Națiunilor Unite. CAEN se aplică în numeroase domenii statistice precum: Registrul statistic al întreprinderilor, Sistemul Conturilor Naționale, în statistica structurală a întreprinderilor, în determinarea indicatorilor industriali pe termen scurt sau în statistica socială, în domeniul forței de muncă.

B. Gruparea datelor statistice

B_1. **Gruparea datelor statistice pe variante** se utilizează, așa cum am arătat anterior, atunci când avem un număr mic de variante și o amplitudine redusă a valorilor variabilei după care se realizează gruparea. În acest caz, se va forma un număr de grupe egal cu numărul de variante. Rezultatele sistematizării se pot prezenta ca în Tabelul 2.2.

Gruparea pe variante a datelor statistice

Tabelul 2.2.

Variantele/valorile caracteristicii de grupare	Număr de unități statistice (n_i)
x_1	n_1
x_2	n_2
...	...
x_i	n_i
...	...
x_r	n_r
Total	$n = \sum_{i=1}^{r} n_i$

unde:

n_i reprezintă numărul unităților care prezintă valoarea x_i a caracteristicii de grupare (se mai numesc frecvențe absolute).

Prin însumarea frecvențelor grupelor (n_i) se obține volumul total al colectivității (n).

Exemplul 2.6. S-au înregistrat numărul de copii pentru 20 de familii (Tabelul 2.3.):

Scanned with OKEN Scanner

Tabelul 2.3.

Familia	Număr de copii
A	1
B	2
C	3
D	0
E	2
F	0
G	1
H	2
I	1
J	3

Familia	Număr de copii
K	2
L	2
M	0
N	2
O	1
P	3
R	1
S	2
T	4
U	2

Variabila „Numărul de copii" are 5 variante. Efectuând o grupare pe variante a familiilor după numărul de copii, se obțin rezultatele din Tabelul 2.4.

Tabelul 2.4.

Număr de copii (x_i)	Număr de familii (n_i)
0	3
1	5
2	8
3	3
4	1
Total	20

B_2. **Gruparea datelor statistice pe intervale de variație** se utilizează atunci când avem un număr mare de variante și o amplitudine mai amplă a valorilor caracteristicii după care se efectuează gruparea.

B_{2_1}. **Gruparea pe intervale egale de variație**

Pentru realizarea grupării pe intervale egale de variație se parcurg următorii pași:

a) se determină amplitudinea variației caracteristicii, ca diferență între valoarea maximă și valoarea minimă a caracteristicii.

$$A = x_{max} - x_{min} \qquad (2.3.)$$

b) se stabilește numărul de grupe. În acest caz pot exista două situații:
- numărul de grupe (r) este prestabilit, pe baza experienței căpătate din studii anterioare asupra domeniului de interes;
- numărul de grupe (r) nu este prestabilit; în acest caz, dacă unitățile se repartizează aproximativ normal după caracteristica studiată, se poate utiliza pentru determinarea numărului de grupe ***relația lui Sturges***:

$$r = 1 + 3{,}322 \cdot lgn \qquad (2.4.)$$

unde n este numărul total de unități ale colectivității.

Scanned with OKEN Scanner

Este recomandat a se folosi un număr potrivit de grupe (de regulă între 4 şi 10). Utilizarea unui număr prea mare de grupe ar duce la fărâmiţarea excesivă a colectivităţii (putând apărea, în acest caz, şi grupe cu frecvenţe nule, iar gruparea ar trebui refăcută); utilizarea, dimpotrivă, a unui număr prea mic de grupe ar putea să nu pună în evidenţă principalele tipuri calitative ale populaţiei după variabila urmărită).

c) se determină mărimea intervalului de grupare (h), ca raport între amplitudinea caracteristicii şi numărul de grupe:

$$h = \frac{A}{r} \qquad (2.5.)$$

Pentru uşurarea calculelor, se recomandă a se folosi mărimi rotunjite de interval, de aceea, dacă valoarea reieşită din calcul este fracţionară, cu mai multe zecimale, ea se poate rotunji prin adaos la o valoare imediat superioară, aleasă în mod convenabil.

De exemplu, dacă $h = 4{,}4225$, se poate rotunji valoarea la $h = 4{,}5$ sau chiar la $h = 5$.

d) se formează intervalele de grupare, prin precizarea limitelor exacte ale acestora.

Limitele intervalelor vor avea acelaşi grad de precizie ca şi datele grupate (acelaşi număr de zecimale), primul interval putând începe chiar de la valoarea minimă a caracteristicii, sau de la o valoare uşor inferioară acesteia, aleasă în mod convenabil. Este bine să nu existe suprapuneri de limite, astfel încât la efectuarea grupării să poată fi respectată condiţia de unicitate. Rezultatele sistematizării pot fi redate cu ajutorul unui tabel (Tabelul 2.5.).

Gruparea datelor statistice pe intervale de variaţie

Tabelul 2.5.

Intervale de variaţie ale caracteristicii de grupare	Număr de unităţi statistice (n_i)
$x_1^{inf} - x_1^{sup}$	n_1
$x_2^{inf} - x_2^{sup}$	n_2
...	...
$x_i^{inf} - x_i^{sup}$	n_i
...	...
$x_r^{inf} - x_r^{sup}$	n_r
Total	$n = \sum_{i=1}^{r} n_i$

Scanned with OKEN Scanner

unde: x_i^{inf} , x_i^{sup} reprezintă limita inferioară, respectiv superioară, a intervalului de variație i.

Dacă $x_i^{sup} \neq x_{i+1}^{inf}$, $i = \overline{1, r-1}$, atunci avem *intervale discrete, discontinue*, întrucât între limita superioară a unui interval şi limita inferioară a intervalului următor există o diferență (*Dif*), numită *unitate de discontinuizare*.

$$x_{i+1}^{inf} = x_i^{sup} + Dif \text{ , pentru } i = \overline{1, r-1} \qquad (2.6.)$$

Dacă $x_i^{sup} = x_{i+1}^{inf}$, $i = \overline{1, r-1}$, atunci avem *intervale continue*, în care limita superioară a unui interval este egală cu limita inferioară a intervalului imediat următor. În acest caz, pentru a respecta condiția de unicitate şi a nu include o unitate statistică în două intervale învecinate, se stabileşte limita care va fi inclusă în interval (limita inferioară sau superioară).

Exemplul 2.7. Se cunoaşte valoarea vânzărilor săptămânale realizate de vânzătorii unui magazin. Valoarea minimă şi valoarea maximă a vânzărilor sunt:

$x_{min} = 1{,}02$ milioane RON;

$x_{max} = 3{,}12$ milioane RON;

Dacă vrem să grupăm datele pe 5 intervale egale, efectuăm următoarele operații:

$A = x_{max} - x_{min} = 3{,}12 - 1{,}02 = 2{,}1$ milioane RON;

$r = 5$

$$h = \frac{A}{r} = \frac{2{,}10}{5} = 0{,}42 \approx 0{,}5 \text{ milioane RON.}$$

Intervalele continue de grupare vor fi:

1,00 – 1,50
1,50 – 2,00
2,00 – 2,50
2,50 – 3,00
3,00 – 3,50

Vom include limita inferioară sau superioară în interval. Putem construi şi intervale discontinue:

1,01 – 1,50	sau	1,00 – 1,49
1,51 – 2,00		1,50 – 1,99
2,01 – 2,50		2,00 – 2,49
2,51 – 3,00		2,50 – 2,99
3,01 – 3,50		3,00 – 3,49

soluții care sunt la fel de valabile.

Scanned with OKEN Scanner

În cazurile în care valoarea minimă şi maximă nu sunt valori aberante, nu sunt foarte depărtate de restul valorilor, atunci ele pot fi incluse în grupare, la primul şi la ultimul interval, construind aceste intervale ca intervale deschise; primul interval nu are precizată limita inferioară, iar ultimul interval nu are stabilită limita superioară.

De exemplu, primul interval (sub 1,50) şi ultimul interval (peste 3,00).

Închiderea intervalelor deschise (valabilă şi când intervalele sunt neegale) se face astfel:

- limita inferioară a primului interval se determină scăzând din limita superioară a primului interval mărimea intervalului următor;
- limita superioară a ultimului interval se determină adunând la limita inferioară a acestuia mărimea intervalului anterior.

Dacă valorile extreme (minim, maxim) sunt şi cazuri aberante, atunci ele pot fi excluse din grupare, fiind studiate separat, dacă unităţile la care s-au înregistrat sunt importante, sau dimpotrivă, pot fi neglijate, dacă nu sunt decât nişte valori accidentale.

B_{2_2}. Gruparea pe intervale neegale de variaţie

Se utilizează atunci când vrem să evidenţiem, în cadrul populaţiei statistice, mai multe tipuri calitative distincte, din punctul de vedere al variabilei studiate. De regulă, în formarea intervalelor neegale se pleacă de la realizarea unei grupări anterioare pe intervale egale, un interval neegal obţinându-se prin comasarea, prin alipirea sau alăturarea mai multor intervale egale.

Exemplul 2.8. Să presupunem că vânzătorii au fost grupaţi după valoarea vânzărilor săptămânale (mil. RON), pe următoarele intervale egale de variaţie:

1,00 – 1,50
1,50 – 2,00
2,00 – 2,50
2,50 – 3,00
3,00 – 3,50
3,50 – 4,00

Prin alipirea intervalelor egale prezentate, se pot obţine următoarele intervale neegale:

1,00 – 2,00 (vânzări cu valoare mică);
2,00 – 2,50 (vânzări cu valoare medie);
2,50 – 4,00 (vânzări cu valoare mare).

Scanned with OKEN Scanner

Scanned with OKEN Scanner

2.2.2. Modalităţi de prezentare şi reprezentare a datelor statistice

2.2.2.1. Serii statistice

În urma sistematizării datelor se constituie ***seriile statistice.***

Seria statistică **reprezintă un mod organizat de prezentare a datelor, sub forma a două şiruri: primul se referă la criteriul de sistematizare, iar al doilea cuprinde datele numerice şi depinde de ordinea de apariţie din primul şir.**

În primul şir pot apărea: variantele caracteristicii, intervale de variaţie ale valorilor numerice sau grupe de variante ale variabilei nenumerice, momente sau intervale de timp, unităţi teritoriale etc. Al doilea şir poate apărea şi sub forma frecvenţelor de apariţie a variantelor din primul şir.

Seriile statistice se pot clasifica după următoarele criterii:

a) în funcţie de conţinutul variabilei după care se face sistematizarea, avem:
   1. serii cronologice (se referă la o variabilă de timp);
   2. serii teritoriale (se referă la o variabilă de spaţiu);
   3. serii de distribuţie de frecvenţe (sau repartiţii de frecvenţe, care se referă la o variabilă atributivă).
b) după natura variabilei, seriile de distribuţie pot fi:
   1. distribuţii heterograde (după o variabilă cantitativă);
   2. distribuţii homograde (după o variabilă calitativă).
c) după numărul variantelor, distribuţiile pot fi:
   1. distribuţii pe variante sau valori;
   2. distribuţii pe grupe de variante (în cazul distribuţiilor homograde) sau pe intervale de valori (în cazul distribuţiilor heterograde).
d) în funcţie de numărul variabilelor după care se face sistematizarea, întâlnim:
   1. distribuţii de frecvenţe unidimensionale (când sistematizarea datelor s-a efectuat după o singură variabilă);
   2. distribuţii de frecvenţe bidimensionale (când sistematizarea datelor s-a efectuat în funcţie de două variabile);
   3. distribuţii multidimensionale (când sistematizarea datelor s-a efectuat în funcţie de trei sau mai multe variabile).

A. SERII DE DISTRIBUŢIE DE FRECVENŢE (REPARTIŢII DE FRECVENŢE)

A.1. Serii de distribuţie de frecvenţe unidimensionale

Seria de distribuţie de frecvenţe unidimensională reprezintă o serie în care primul şir cuprinde variantele/valorile sau intervalele de variaţie ale unei variabile, iar al doilea şir – frecvenţele de apariţie ale variantelor sau volumul grupelor.

A.1.1. Distribuţii heterograde

Se pot reprezenta sub forma:

– pentru o caracteristică discretă (repartiţii de frecvenţe pe variante/valori):

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_r \\ n_1 & n_2 & \dots & n_i & \dots & n_r \end{pmatrix}, \tag{2.7.}$$

unde n_i $(i=\overline{1,r})$ sunt frecvenţele de apariţie ale variantei x_i.

– pentru o caracteristică continuă (repartiţie de frecvenţe pe intervale de valori):

$$X:\begin{pmatrix} x_1^{\inf} < X < x_1^{\sup} & x_2^{\inf} < X < x_2^{\sup} & \dots & x_i^{\inf} < X < x_i^{\sup} & \dots & x_r^{\inf} < X < x_r^{\sup} \\ n_1 & n_2 & \dots & n_i & \dots & n_r \end{pmatrix} \tag{2.8.}$$

sau

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_r \\ n_1 & n_2 & \dots & n_i & \dots & n_r \end{pmatrix} \tag{2.9.}$$

unde x_i, $i=\overline{1,r}$ sunt centrele intervalelor de variaţie.

Distribuţiile prezentate sunt *distribuţii empirice*, deoarece ele conţin date reale, observate din realitatea economico-socială. Modelul teoretic al distribuţiilor empirice îl reprezintă distribuţiile (repartiţii) teoretice, numite şi *distribuţii de probabilitate* şi care sunt de forma:

– pentru o variabilă discretă (pe variante):

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_r \\ p_1 & p_2 & \dots & p_i & \dots & p_r \end{pmatrix}, \tag{2.10.}$$

unde p_i sunt probabilităţile de apariţie ale variantei/valorii x_i a caracteristicii X (sau probabilitatea ca variabila aleatorie X să ia valoarea x_i).

– pentru o variabilă continuă (pe intervale de variaţie) (funcţia densităţii de probabilitate):

$$X:\begin{pmatrix} X < x_1 & x_1 \le X \le x_2 & \dots & x_{i-1} \le X \le x_i & \dots & x_r \le X \\ f(x_1) & f(x_2) & \dots & f(x_i) & \dots & f(x_{r+1}) \end{pmatrix} \tag{2.11.}$$

unde: $f(x_i) = P(x_{i-1} \le X \le x_i),\ i=\overline{2,r}$ (2.12.)

$$f(x_1) = P(X \le x_1) \tag{2.13.}$$

$$f(x_{r+1}) = P(x_r \le X) \tag{2.14.}$$

Scanned with OKEN Scanner

➢ **Serii de repartiție de frecvențe pe variante (discrete)**

Forma generală a unei serii de repartiție de frecvențe pe variante se prezintă în Tabelul 2.2. Se pot întâlni următoarele tipuri de frecvențe:

- **frecvențe absolute**

Frecvența absolută a unei grupe este dată de numărul de unități care se încadrează în acea grupă (sau care au valoarea x_i a caracteristicii). Frecvențele absolute se notează cu n_i $(i = \overline{1,r})$. Distribuția din Tabelul 2.2. este o distribuție de frecvențe absolute.

- **frecvențe relative**

Frecvența relativă a unei grupe arată ce pondere din numărul total de unități se regăsește în acea grupă și se calculează ca raport între frecvența absolută a grupei și volumul total al populației. Frecvențele relative se notează cu n_i^* $(i = \overline{1,r})$ și se determină cu următoarea relație:

$$n_i^* = \frac{n_i}{\sum_{i=1}^{r} n_i} = \frac{n_i}{n} \text{ și se exprimă în coeficienți,} \qquad (2.15.)$$

sau

$$n_{i\%}^* = \frac{n_i}{\sum_{i=1}^{r} n_i} \cdot 100 = \frac{n_i}{n} \cdot 100 \text{ și se exprimă în procente.} \qquad (2.16.)$$

O distribuție de frecvențe relative are forma generală prezentată în Tabelul 2.6.

Distribuția de frecvențe relative, pe variante

Tabelul 2.6.

Variantele/valorile caracteristicii de grupare	Frecvențe relative (n_i^*)
x_1	n_1^*
x_2	n_2^*
...	...
x_i	n_i^*
...	...
x_r	n_r^*
Total	$\sum_{i=1}^{r} n_i^* = 1$

Scanned with OKEN Scanner

Suma frecvențelor relative este 1 sau 100, după cum sunt exprimate în coeficienți sau în procente.

$$\sum_{i=1}^{r} n_i^* = 1 \qquad \text{sau} \qquad \sum_{i=1}^{r} n_{i\%}^* = 100 \tag{2.17.}$$

- **frecvențe cumulate**

Se pot cumula atât frecvențele absolute, cât și cele relative; cumularea se poate efectua crescător sau descrescător.

– ***frecvența absolută cumulată crescător*** *a unei grupe (i) reprezintă numărul de unități care au valoarea caracteristicii mai mică (sau cel mult egală) cu varianta corespunzătoare grupei* (x_i).

$$Fc_i = \sum_{k=1}^{i} n_k \tag{2.18.}$$

Frecvența absolută cumulată crescător a ultimei grupe este egală cu volumul colectivității (cu n):

$$Fc_r = \sum_{k=1}^{r} n_k = n \tag{2.19.}$$

– ***frecvența relativă cumulată crescător*** *a unei grupe (i) este egală cu ponderea unităților statistice care au valoarea caracteristicii mai mică (sau cel mult egală) cu varianta corespunzătoare grupei* (x_i), *în numărul total de unități.*

$$Fc_i^* = \sum_{k=1}^{i} n_k^* \tag{2.20.}$$

Frecvența relativă cumulată crescător a ultimei grupe este egală cu suma tuturor frecvențelor relative, adică 1 sau 100.

$$Fc_r^* = \sum_{k=1}^{r} n_k^* = 1 \text{ sau } 100 \tag{2.21.}$$

– ***frecvența absolută cumulată descrescător*** *a unei grupe (i) este egală cu numărul unităților statistice care au valoarea caracteristicii mai mare (sau eventual egală) cu varianta corespunzătoare grupei* (x_i).

$$Fd_i = \sum_{k=i}^{r} n_k \tag{2.22.}$$

Frecvența absolută cumulată descrescător a primei grupe este egală cu numărul total de unități statistice (cu n):

Scanned with OKEN Scanner

$$Fd_1 = \sum_{k=1}^{r} n_k = n \qquad (2.23.)$$

– ***frecvența relativă cumulată descrescător*** *a unei grupe (i) este egală cu ponderea unităților statistice care au valoarea caracteristicii mai mare (sau eventual egală) cu varianta corespunzătoare grupei (x_i), în numărul total de unități.*

$$Fd_i^* = \sum_{k=i}^{r} n_k^* \qquad (2.24.)$$

Frecvența relativă cumulată descrescător a primei grupe este egală cu suma tuturor frecvențelor relative, adică 1 sau 100.

$$Fd_1^* = \sum_{k=1}^{r} n_k^* = 1 \text{ sau } 100 \qquad (2.25.)$$

Între frecvența absolută cumulată crescător şi frecvența absolută cumulată descrescător a unei grupe există următoarea relație:

$$Fc_i + Fd_i = n + n_i \qquad (2.26.)$$

Exemplul 2.9. Considerăm distribuția de frecvențe din Tabelul 2.4., pentru care s-au calculat frecvențele relative, frecvențele absolute şi relative cumulate crescător şi descrescător (Tabelul 2.7.):

Tabelul 2.7.

Număr de copii (x_i)	Număr de familii (n_i)	Frecvențe relative (n_i^*)	Frecvențe absolute cumulate		Frecvențe relative cumulate	
			Crescător	Descres-cător	Crescător	Descres-cător
0	1	2	3	4	5	6
0	3	0,15	3	20	0,15	1,00
1	5	0,25	8	17	0,40	0,85
2	8	0,40	16	12	0,80	0,60
3	3	0,15	19	4	0,95	0,20
4	1	0,05	20	1	1,00	0,05
Total	20	1,00	–	–	–	–

➢ **Serii de repartiție de frecvențe pe intervale de valori (continue)**

Forma generală a unei astfel de serii este prezentată în Tabelul 2.5.

- ***Mărimea intervalului de variație*** se calculează cu una din următoarele relații:

a) $h_i = x_i^{sup} - x_i^{inf} + Dif \;,\; i = \overline{1,r}$ (2.27.)

Scanned with OKEN Scanner

În cazul intervalelor continue, $Dif = 0$ și deci relația devine:

$$h_i = x_i^{sup} - x_i^{inf}, \; i = \overline{1,r} \tag{2.28.}$$

b) $$h_i = x_{i+1}^{inf} - x_i^{inf}, \; i = \overline{1,r} \tag{2.29.}$$

sau

$$h_i = x_i^{sup} - x_{i-1}^{sup}, \; i = \overline{2,r} \tag{2.30.}$$

- ***Centrul intervalului*** este valoarea caracteristicii întâlnită la fiecare unitate din intervalul respectiv, (luându-se în considerare că valorile variabilei urmează o repartiție aproximativ normală în intervalul respectiv). Se determină diferit, în funcție de tipul variației:
  - pentru intervalele continue, centrul intervalului se calculează ca medie aritmetică simplă a limitelor intervalului:

$$x_i = \frac{x_i^{inf} + x_i^{sup}}{2} \tag{2.31.}$$

  - pentru intervalele discontinue, centrul intervalului se determină adunând la limita inferioară a acestuia jumătate din mărimea intervalului:

$$x_i = x_i^{inf} + \frac{h_i}{2} \tag{2.32.}$$

Relația (2.32.) este valabilă și în cazul intervalelor continue.

- ***Frecvența absolută a grupei*** (n_i) este egală cu numărul de unități statistice care au valoarea caracteristicii mai mare (sau egală) cu limita inferioară a intervalului și mai mică (sau egală) cu limita superioară a acesteia.
- ***Frecvența relativă a unei grupe*** (n_i^*) reprezintă ponderea unităților statistice în volumul total al colectivității care au valoarea caracteristicii cuprinsă între limita inferioară și cea superioară a grupei respective.
- ***Frecvențele cumulate***:

– *frecvența absolută cumulată crescător* a unei grupe este egală cu numărul unităților care au valoarea variabilei mai mică (sau egală) cu limita superioară a grupei (mai exact între x_1^{inf} și x_i^{sup});

– *frecvența relativă cumulată crescător* a unei grupe este egală cu ponderea unităților în numărul total al acestora, care au valoarea variabilei mai mică (sau egală) cu limita superioară a grupei;

– *frecvența absolută cumulată descrescător* a unei grupe este egală cu numărul unităților pentru care valoarea caracteristicii este mai mare (sau egală) cu limita inferioară a grupei (mai exact între x_i^{inf} și x_r^{sup});

Scanned with OKEN Scanner

– *frecvența relativă cumulată descrescător* a unei grupe este egală cu ponderea unităților în numărul total al acestora, care au valoarea caracteristicii mai mare (sau egală) cu limita inferioară a grupei.

Frecvențele relative, precum și frecvențele cumulate se calculează pe baza relațiilor (2.15.)-(2.25.)

Exemplul 2.10. Presupunem că avem următoarea distribuție a vânzătorilor unui magazin în funcție de valoarea vânzărilor săptămânale realizate (Tabelul 2.8., coloanele 0, 1):

Tabelul 2.8.

Intervale de variație a vânzărilor (mil. RON)	Nr. de vânzători (n_i)	Frecvențe relative (n_i^*) (%)	Frecvențe absolute cumulate		Frecvențe relative cumulate (%)		Centre de interval (x_i)
			Crescător	Descrescător	Crescător	Descrescător	
0	1	2	3	4	5	6	7
1,00-1,50	10	5	10	200	5	100	1,25
1,50-2,00	20	10	30	190	15	95	1,75
2,0-2,50	120	60	150	170	75	85	2,25
2,50-3,00	30	15	180	50	90	25	2,75
3,00-3,50	20	10	200	20	100	10	3,25
Total	200	100	–	–	–	–	–

Pentru această distribuție s-au calculat frecvențele relative, frecvențele absolute și relative cumulate crescător și descrescător, precum și contrele de interval (Tabelul 2.8., coloanele 2-7).

În cazul distribuțiilor heterograde *pe intervale neegale de variație*, nu se lucrează cu frecvențele absolute obișnuite, ci cu *frecvențe reduse* la un interval etalon (*frecvențe corectate*). Se alege un interval de variație ca fiind intervalul etalon (cu mărimea h_{et}) și se determină *coeficienți de corecție* prin raportarea mărimii fiecărui interval la mărimea intervalului etalon:

$$k_i^{cor} = \frac{h_i}{h_{et}} \qquad (2.33.)$$

De obicei se alege ca interval etalon intervalul cu mărimea cea mai mică, astfel încât coeficienții de corecție determinați să aibă valori supraunitare. Frecvența corectată a unui interval se determină împărțind frecvența absolută a acestuia la coeficienții de corecție:

$$n_i^{cor} = \frac{n_i}{k_i^{cor}} \qquad (2.34.)$$

Exemplul 2.11. Să considerăm distribuția pe intervale neegale a vânzătorilor după valoarea vânzărilor săptămânale (mil. RON) (Tabelul 2.9. coloanele 0, 1):

Scanned with OKEN Scanner

Tabelul 2.9.

Intervale de variație a vânzărilor (mil. RON)	Nr. de vânzători (n_i)	Mărimea intervalului (h_i)	Coeficienți de corecție (k_i^{cor})	Frecvențe reduse (corectate) (n_i^{cor})
0	1	2	3	4
1,00 – 2,00	30	1,00	2	15,00
2,00 – 2,50	120	0,50	1	120,00
2,50 – 4,00	50	1,50	3	16,67
Total	200	–		–

A.1.2. Distribuții homograde

Și pentru distribuțiile homograde, în care sistematizarea datelor este realizată după o variabilă calitativă, se pot calcula frecvențe absolute și relative.

În cazul acestor distribuții, populația este structurată pe mai multe tipuri calitative diferite (sau pe categorii), frecvența absolută reprezentând numărul unităților care se încadrează într-o categorie, iar frecvența relativă – ponderea unităților de un anumit tip calitativ, în volumul total al colectivității.

Variantele caracteristicii după care s-a efectuat sistematizarea datelor pot fi măsurate pe o scală nominală sau ordinală. Numai în acest din urmă caz are sens determinarea frecvențelor cumulate (absolute și relative).

Exemplul 2.12. Distribuția absolvenților unei facultăți economice după domeniul în care s-au angajat este (Tabelul 2.10. – coloanele 0, 1):

Tabelul 2.10.

Domeniu	Număr de absolvenți (n_i)	Ponderea absolvenților (n_i^*) (%)
0	1	2
Contabilitate	95	31,7
Marketing	72	24,0
Finanțe	55	18,3
Management economic	43	14,3
Altele sau fără loc de muncă	35	11,7
Total	300	100,0

Distribuția din Tabelul 2.10. este o distribuție homogradă după o variabilă măsurată pe scală nominală, pentru care s-au determinat frecvențele relative (Tabelul 2.10., coloana 2).

Exemplul 2.13. Distribuția studenților unei grupe după calificativul obținut la un proiect este (Tabelul 2.11., coloanele 0, 1):

Scanned with OKEN Scanner

Tabelul 2.11.

Calificativ (x_i)	Număr de studenți (n_i)	Frecvențe relative (n_i^*)	Frecvențe absolute cumulate		Frecvențe relative cumulate	
			Crescător	Descres-cător	Crescător	Descres-cător
0	1	2	3	4	5	6
Insuficient	3	0,10	3	30	0,10	1,00
Satisfăcător	4	0,13	7	27	0,23	0,90
Bine	15	0,50	22	23	0,73	0,77
Foarte bine	6	0,20	28	8	0,93	0,27
Excelent	2	0,07	30	2	1,00	0,07
Total	30	1,00	–	–	–	–

Distribuția din Tabelul 2.11. este o distribuție homogradă după o variabilă măsurată pe scală ordinală, pentru care s-au calculat frecvențele relative și cele cumulate (Tabelul 2.11., coloanele 2-6).

A.2. Serii de distribuție de frecvențe bidimensionale

Forma generală a unei distribuții de frecvențe bidimensionale, în care s-au luat în considerare două variabile statistice X și Y este prezentată în Tabelul 2.12.

Distribuția de frecvențe bidimensională

Tabelul 2.12.

Variante sau centre de interval pt. variabila X	Variante sau centre de interval pt. variabila Y						Total unități
	y_1	y_2	...	y_j	...	y_p	
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
...	...	...	...	...	...	...	...
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rp}	$n_{r.}$
Total unități	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.p}$	$n_{..}$

Tabelul 2.12. reprezintă un ***tabel de corelație***, în care avem:

x_i, $i=\overline{1,r}$ este varianta sau centrul de interval pentru grupa i, formată după valorile variabilei X;

y_j, $j=\overline{1,p}$ este varianta sau centrul de interval al grupei j, formată după valorile variabilei Y;

n_{ij}, $i=\overline{1,r}$; $j=\overline{1,p}$ reprezintă numărul unităților statistice la care întâlnim simultan valoarea x_i a caracteristicii X și valoarea y_j a caracteristicii Y;

Scanned with OKEN Scanner

$n_{i.} = \sum_{j=1}^{p} n_{ij}$ este numărul de unități statistice care au valoarea x_i a caracteristicii X, indiferent de valoarea caracteristicii Y;

$n_{.j} = \sum_{i=1}^{r} n_{ij}$ este numărul de unități statistice care au valoarea y_j a caracteristicii Y, indiferent de valoarea caracteristicii X;

$n_{..} = n = \sum_{i=1}^{r}\sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{p} n_{.j}$ este volumul total al colectivității.

În tabelul de corelație distingem următoarele distribuții:

– o distribuție bidimensională, de forma: $\begin{Bmatrix} x_i, y_j \\ n_{ij} \end{Bmatrix}$, $i = \overline{1,r}$; $j = \overline{1,p}$

– două distribuții marginale, de forma:

- după X: $\begin{Bmatrix} x_i \\ n_{i.} \end{Bmatrix}$, $i = \overline{1,r}$
- după Y: $\begin{Bmatrix} y_j \\ n_{.j} \end{Bmatrix}$, $j = \overline{1,p}$

– $r + p$ distribuții condiționate, de forma:

- pentru fiecare $X = x_i$, avem câte o distribuție după Y: $\begin{Bmatrix} y_j \\ n_{ij} \end{Bmatrix}$, $j = \overline{1,p}$
- pentru fiecare $Y = y_j$, avem câte o distribuție după X: $\begin{Bmatrix} x_i \\ n_{ij} \end{Bmatrix}$, $i = \overline{1,r}$

Dacă cele două variabile condiționate sunt caracteristici alternative (sau care se pot transforma în caracteristici alternative), se alcătuiește ***tabelul de asociere***, de forma (Tabelul 2.13.):

Tabelul de asociere

Tabelul 2.13.

Clase după X	Clase după Y		Total unități
	Y (y_1)	Non Y (y_2)	
X (x_1)	n_{11}	n_{12}	$n_{1.} = n_{11} + n_{12}$
Non X (x_2)	n_{21}	n_{22}	$n_{2.} = n_{21} + n_{22}$
Total unități	$n_{.1} = n_{11} + n_{21}$	$n_{.2} = n_{12} + n_{22}$	$n_{..} =$

Scanned with OKEN Scanner

Exemplul 2.14. Pentru 20 de instalaţii industriale ale unui agent economic s-au înregistrat următorele date (Tabelul 2.14.):

Tabelul 2.14.

Instalaţia	Durata de funcţionare (ani)	Costuri de întreţinere (unit. monetare)	Instalaţia	Durata de funcţionare (ani)	Costuri de întreţinere (unit. monetare)
X1	6	92	X11	3	110
X2	7	181	X12	8	105
X3	1	23	X13	9	181
X4	3	40	X14	8	152
X5	6	126	X15	6	100
X6	4	82	X16	3	45
X7	5	117	X17	5	120
X8	2	49	X18	8	107
X9	1	64	X19	7	178
X10	9	141	X20	2	50

Se observă că în Tabelul 2.14. sunt redate variantele (negrupate) ale caracteristicilor *X* şi *Y*, unde *X* este durata de funcţionare, iar *Y* este costul de întreţinere.

Exemplul 2.15. Pentru 24 de salariaţi care lucrează în sectorul vânzărilor, s-au înregistrat vechimea în muncă (ani) şi valoarea vânzărilor (unităţi monetare – u.m.). Datele sunt prezentate în Tabelul 2.15.:

Tabelul 2.15.

Grupe de salariaţi după valoarea vânzărilor (u.m.)	Grupe de salariaţi după vechime (ani)					Total salariaţi
	0–5	5–10	10–15	15–20	20–25	
160 – 180	–	–	–	5	1	6
140 – 160	–	1	7	–	–	8
120 – 140	–	6	–	-	–	6
100 – 120	4	–	–	–	–	4
Total salariaţi	4	7	7	5	1	24

Datele din Tabelul 2.15. (tabel de corelaţie) sunt grupate pe intervale de variaţie, după valorile ambelor variabile.

Alte interpretări pe baza Tabelelor 2.12. şi 2.13. vor fi detaliate în capitolul 5.

B. SERII CRONOLOGICE

Seria cronologică **reprezintă un tip de serie statistică în care criteriul de ordonare, de sistematizare a datelor este unul temporal: timpul, fie că e vorba de momente sau de intervale de timp.**

Scanned with OKEN Scanner

Forma generală a unei serii cronologice este:

$$\begin{pmatrix} t_1 & t_2 & \dots & t_i & \dots & t_n \\ y_{t1} & y_{t2} & \dots & y_{ti} & \dots & y_{tn} \end{pmatrix} = \begin{Bmatrix} t_i \\ y_{ti} \end{Bmatrix}, \; i = \overline{1,n} \qquad (2.35.)$$

unde $t_i \; (i = \overline{1,n})$ este momentul sau intervalul de timp i;

$y_{ti} \; (i = \overline{1,n})$ este nivelul indicatorului la momentul sau intervalul de timp i.

Exemplul 2.16; Evoluția ratei șomajului în România în perioada 1990-2003 este redată în Tabelul 2.16.:

Evoluția ratei șomajului din România în perioada 1990-2003

Tabelul 2.16.

Ani	Rata șomajului (%)
1990	0
1991	3,0
1992	8,2
1993	10,4
1994	10,9
1995	9,5
1996	6,7
1997	6,0
1998	6,3
1999	6,8
2000	7,1
2001	6,6
2002	8,4
2003	7,0

SURSA: Anuarul statistic al României, 2002, 2003, 2004, INS.

Analiza statistică a seriilor cronologice va fi detaliată în capitolul 6.

C. SERII TERITORIALE

***Seria teritorială* este rezultatul sistematizării datelor statistice după o variabilă de spațiu (teritorială).**

Are următoarea formă generală:

$$\begin{pmatrix} ut_1 & ut_2 & \dots & ut_i & \dots & ut_n \\ y_{ut1} & y_{ut2} & \dots & y_{uti} & \dots & y_{utn} \end{pmatrix} = \begin{Bmatrix} ut_i \\ y_{ti} \end{Bmatrix}, \; i = \overline{1,n} \qquad (2.36.)$$

unde $ut_i \; (i = \overline{1,n})$ este unitatea teritorială i;

$y_{uti} \; (i = \overline{1,n})$ este nivelul indicatorului pentru unitatea teritorială i

Scanned with OKEN Scanner

Exemplul 2.17. Produsul Intern Brut pe locuitor, pentru unele țări membre ale Uniunii Europene, în anul 2002 a fost (Tabelul 2.17.):

Produsului Intern Brut pe locuitor, pentru unele țări membre ale Uniunii Europene, în anul 2002

Tabelul 2.17.

Țara	PIB/loc. în 2002 (mii EURO/pers.)
Belgia	24,95
Danemarca	33,52
Germania	25,04
Grecia	11,93
Spania	16,13
Franța	23,99
Irlanda	30,53
Italia	21,02
Luxemburg	55,0
Olanda	26,69
Austria	26,05
Portugalia	12,18
Finlanda	25,96
Suedia	26,29
Marea Britanie	26,56

SURSA: EUROSTAT.

2.2.2.2. Grafice şi tabele statistice

Graficele şi tabelele statistice, ca modalități de prezentare şi reprezentare a datelor statistice, alături de serii statistice, ne ajută, adesea, să înțelegem mai uşor şi mai bine rezultatele unei investigații statistice. Ele sunt părți esențiale ale analizei datelor, oferind modalități de ordonare, de organizare şi de interpretare a datelor statistice. Deseori, un grafic sau un tabel bine întocmit „spune" mai mult decât o descriere în cuvinte a rezultatelor cercetării sau, în orice caz, ne comunică mai repede mesajul. Pentru a parcurge un text sunt necesare, de exemplu, câteva minute, dar pentru a „citi" un grafic sunt de ajuns doar câteva secunde. Totodată, graficele corect elaborate au darul de a menține mesajul comunicat mai mult şi mai uşor în memorie, decât un text cursiv.

Desigur, nu trebuie să înțelegem că orice tabel sau grafic incluse în raportul cercetării statistice ar îmbunătăți neapărat calitatea acestuia. Un grafic întocmit necorespunzător poate distorsiona rezultatul real al investigației, sau ar putea crea confuzie. Este necesar să ştim să facem diferența între graficele de popularizare

Scanned with OKEN Scanner

care apar adesea în presă, la televizor sau în diverse afişe şi graficele ştiinţifice, care se întocmesc cu respectarea riguroasă a unor reguli.

Graficele şi tabelele statistice sunt modalităţi de reprezentare într-un mod sugestiv a datelor colectate, prelucrate şi rezultate dintr-un studiu statistic, rezumând cele mai importante caracteristici ale acestora.

Oportunitatea utilizării tabelelor sau graficelor, ca modalitate de reprezentare a datelor depinde, adesea, de scopul principal urmărit; dacă acesta este de a ilustra, de exemplu, relaţia dintre diferite grupuri sau de a urmări un anumit trend, atunci este mai indicat să utilizăm un grafic; dacă, însă, scopul de bază este prezentarea exactă a valorilor numerice ale datelor, atunci este mai oportun să folosim un tabel.

Grafice statistice

În prezent există o mulţime de programe pe calculator care permit, cu tot mai mare uşurinţă, construirea unor grafice din cele mai variate tipuri, de la cele mai simple la cele foarte complexe. Trebuie, desigur, ţinut seama de opţiunile pe care le oferă aceste programe, de aria de posibilităţi pe care o dezvoltă, însă, de multe ori, este mai indicat să utilizăm o serie de reprezentări mai simple în locul figurilor prea complexe şi complicate, care, la un moment dat, pot ascunde sau chiar altera, distorsiona mesajul pe care vrem să-l transmitem.

Alegerea tipului de grafic depinde de tipologia datelor pe care vrem să le reprezentăm (dacă sunt continue sau discrete). Astfel, putem utiliza grafice într-un sistem de coordonate, grafice cu ajutorul unor figuri geometrice, grafice cu ajutorul hărţior sau al altor figuri naturale sau simbolice.

Pentru ca un grafic să-şi atingă scopul pentru care a fost proiectat, el trebuie să fie corect întocmit, să respecte anumite reguli generale, să cuprindă o serie de elemente obligatorii sau opţionale şi anume:

- *titlul graficului* – este un element obligatoriu al oricărui grafic, care trebuie să reflecte clar fenomenul studiat, cu precizarea coordonatelor temporale, spaţiale şi organizatorice;
- *sistemul de coordonate* utilizat la trasarea graficelor; principalele tipuri de sisteme de coordonate sunt:
  - rectangulare (ortogonale) – care sunt întâlnite în majoritatea graficelor; sistemul este format din două axe perpendiculare, una orizontală (axa absciselor) Ox şi alta verticală (axa ordonatelor) Oy; pe axa absciselor se reprezintă, de regulă, valorile variabilei independente, iar axa ordonatelor este folosită pentru a reprezenta valorile variabilei dependente; variabila independentă este utilizată pentru a previziona, a descrie sau a explica diferenţele sau modificările survenite în variabila dependentă, ultima reprezentând variabila studiată prin investigaţia

Scanned with OKEN Scanner

statistică. Un punct din cadranul delimitat de cele două axe are coordonatele (x_i, y_i). Specialiştii consideră că cel mai indicat din punct de vedere estetic este graficul sub forma unui dreptunghi, a cărui lungime este de 1,6 ori mai mare decât lăţimea sa.

Pe cele două axe trebuie specificată în clar denumirea variabilelor care se reprezintă cu ajutorul lor şi scările de reprezentare.

- polare (radiale) – sunt mai rar utilizate; se folosesc în cazul reprezentării unor serii cronologice afectate de factori sezonieri. Un punct din reţeaua polară este definit prin distanţa de la acel punct până în centrul cercului (raza) şi prin unghiul dinte acea rază şi raza orizontală.

- *reţeaua graficului* este un element opţional, al cărui rol este de a facilita reprezentarea datelor în cadrul graficului; reţeaua este formată, în funcţie de sistemul de coordonate folosit, dintr-un complex de linii paralele, orizontale şi verticale, trasate mai subţire (în cazul sistemului de coordonate ortogonale) şi printr-un sistem de cercuri concentrice (în cazul utilizării unui sistem de coordonate polare).
- *scara de reprezentare*, element obligatoriu, arată ce unitate din valoarea variabilei se reprezintă printr-o unitate de lungime aleasă drept etalon pe grafic. Scara se alege în funcţie de nivelul valoric al datelor de reprezentat, de scopul urmărit, de spaţiul pe care-l avem disponibil pentru trasarea graficului. Scările pot fi:
  - aritmetice (în care atât pe axa Ox, cât şi pe Oy, distanţele sunt proporţionale cu valorile variabilelor);
  - logaritmice (în care pe ambele axe distanţele sunt proporţionale cu logaritmii valorilor variabilelor);
  - semilogaritmice (când pe axa Ox s-a folosit o scară aritmetică, iar pe Oy o scară logaritmică).
- *legenda* este un element opţional, care explică semnificaţia simbolurilor utilizate în cadrul graficului.

Pe lângă aceste elemente menţionate, graficul mai poate cuprinde şi note explicative, sursa datelor etc.

Principalele tipuri de reprezentări grafice sunt:

A. Graficul (diagrama) prin coloane

Este folosit pentru reprezentarea distribuţiilor de frecvenţe absolute sau relative, în care sistematizarea s-a făcut după o variabilă categorială, calitativă, măsurată pe scală nominală. Graficul se trasează în sistemul de axe rectangulare Ox şi Oy, pe Ox se reprezintă categoriile variabilei calitative, iar pe Oy frecvenţele (absolute sau relative) sau nivelul indicatorului. Aşadar, graficul constă dintr-o succesiune de coloane de lăţime egală, câte o coloană pentru fiecare catego-

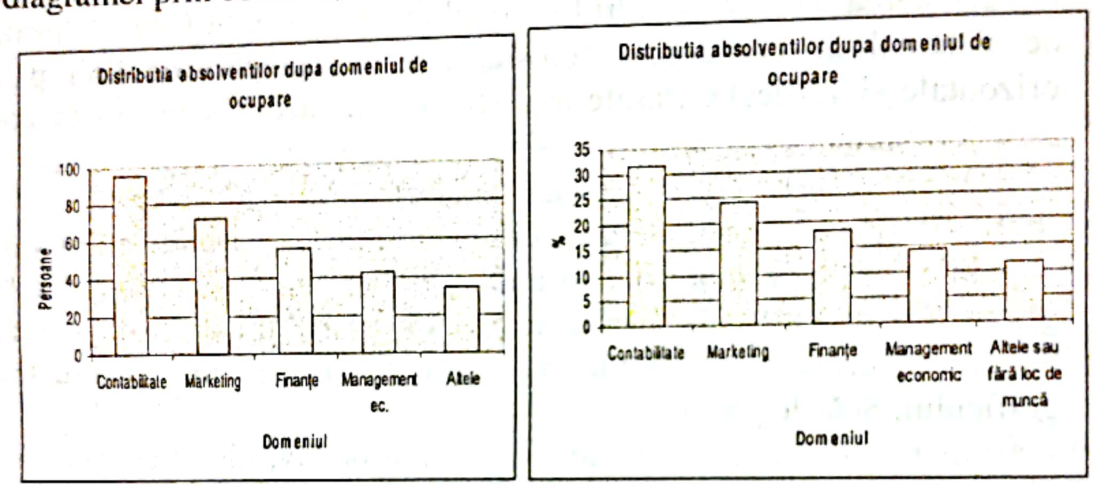
Scanned with OKEN Scanner

rie/variantă a variabilei nominale, egal distanțate între ele (la distanțe, de regulă, mai mici decât grosimea coloanelor) și cu înălțimea proporțională cu frecvențele sau nivelul indicatorului corespunzător categoriei respective.

Este important ca acest coloane să nu fie lipite unele de altele, deoarece nu trebuie să dea senzația de continuitate a categoriilor de pe axa Ox. Denumirea variantei/categoriei reprezentate printr-o coloană se trece la mijlocul acesteia.

Graficul are avantajul că este relativ ușor de construit (fiind unul din cele mai larg utilizate tipuri de grafice) și suficient de flexibil, astfel încât să permită rearanjarea, reordonarea coloanelor pentru efectuarea unor comparații.

Exemplul 2.18. Datele din Tabelul 2.10. se pot reprezenta grafic cu ajutorul diagramei prin coloane, astfel (Figura 2.1.):



a) distribuția de frecvențe absolute b) distribuția de frecvențe relative

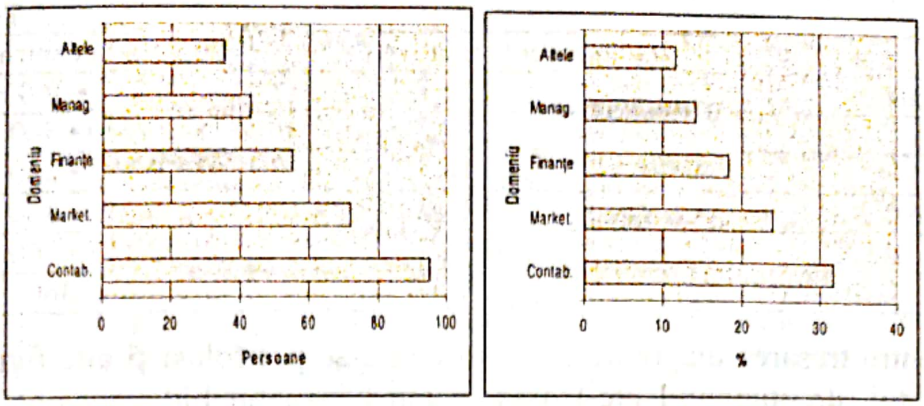
Figura 2.1. Grafice prin coloane

B. Graficul (diagrama) prin benzi (sau bare)

Graficul se construiește asemănător cu o diagramă prin coloane, în care, practic, se inversează cele două axe: pe axa Ox se reprezintă frecvențele absolute sau relative sau nivelul indicatorului, iar pe axa Oy se reprezintă categoriile variabilei calitative. Benzile arată, deci, ca niște coloane culcate pe orizontală, care au bazele pe axa Oy, grosime egală, sunt egal distanțate unele de altele și au lungimea proporțională cu mărimea frecvențelor sau cu nivelul indicatorului. Și coloanele și benzile se pot ordona descrescător după mărimea frecvențelor sau după nivelul indicatorului.

Exemplul 2.19. Datele din Tabelul 2.10. pot fi reprezentate grafic și sub formă de benzi (Figura 2.2.):

Scanned with OKEN Scanner



a) distribuția de frecvențe absolute b) distribuția de frecvențe relative

Figura 2.2. Distribuția absolvenților după domeniul de ocupare

C. Diagrama de structură

Diagrama de structură este folosită pentru a reprezenta structura unei colectivități, sistematizate după valorile unei variabile cantitative sau calitative. Graficul arată, aşadar, modul în care întregul se subdivide în părți componente. De obicei, diagrama se trasează cu ajutorul cercului, a cărui arie reprezintă întregul; acesta se împarte în mai multe bucăți („felii"), aria unei „felii" reprezentând frecvența relativă a unei clase. Unghiul la centru corespunzător acestei părți de cerc este proporțional cu raportul dintre frecvența absolută şi volumul total al colectivității (adică cu frecvența relativă) a acelei clase/grupe.

Exemplul 2.20. Dacă 30% din unitățile populației se încadrează într-o grupă, atunci unghiul la centru corespunzător „feliei" acelei grupe va fi de: $360 \cdot 0{,}3 = 108°$.

Pentru a face deosebirea între părțile corespunzătoare claselor/grupelor în care este structurată colectivitatea, ele se haşurează sau se colorează diferit, precizându-se acest sistem de diferențiere într-o legendă.

Exemplul 2.21. Datele din Tabelul 2.10., pentru care s-au calculat unghiurile la centru corespunzătoare fiecărui domeniu de angajare (Tabelul 2.18., s-au reprezentat grafic cu ajutorul diagramei de structură în Figura 2.3.

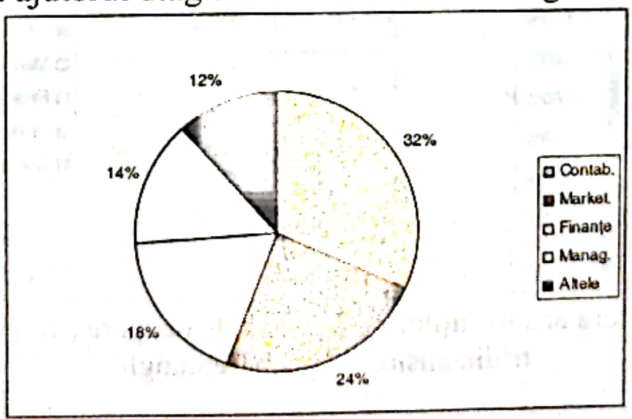


Figura 2.3. Structura absolvenților după domeniul de ocupare

Scanned with OKEN Scanner

Tabelul 2.18.

Domeniul de angajare	Frecvenţe relative (%)	Unghiul la centru (°)
Contabilitate	31,7	0,317 • 360 = 114
Marketing	24,0	0,240 • 360 = 86
Finanţe	18,3	0,183 • 360 = 66
Management ec.	14,3	0,143 • 360 = 51
Altele sau fără loc de muncă	11,7	0,117 • 360 = 43
Total	100,0	360

Pentru trasarea diagramelor de structură se pot folosi şi alte figuri geometrice (pătratul, dreptunghiul etc.), trasate într-un spaţiu bidimensional sau chiar tridimensional. Adesea, reprezentarea tridimensională nu redă cu atâta acurateţe proporţiile dintre părţile întregului, estompând rezultatul final, aşa cum se observă şi din Figurile 2.4., 2.5., 2.6., 2.7.

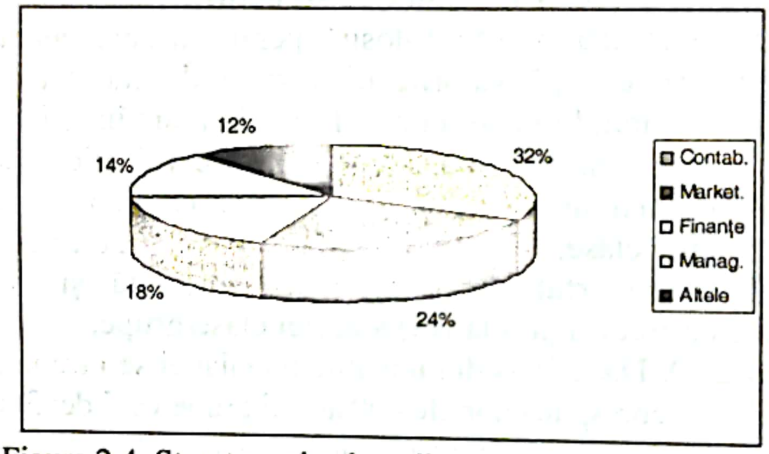


Figura 2.4. Structura absolvenţilor pe domenii de ocupare (diagramă de structură tridimensională prin cerc)

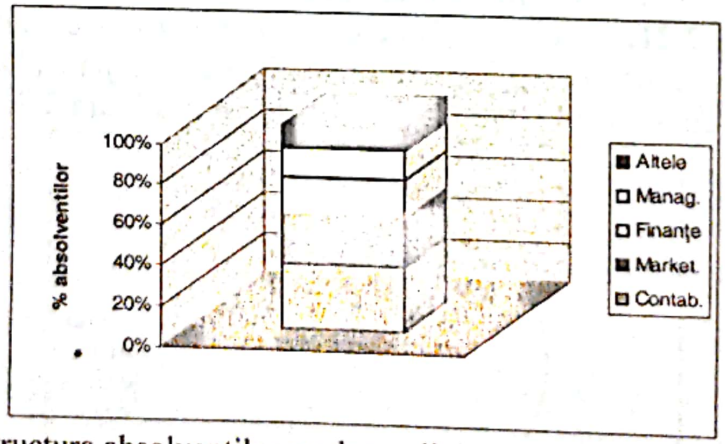


Figura 2.5. Structura absolvenţilor pe domenii de ocupare (diagramă de structură tridimensională prin dreptunghi)

Scanned with OKEN Scanner

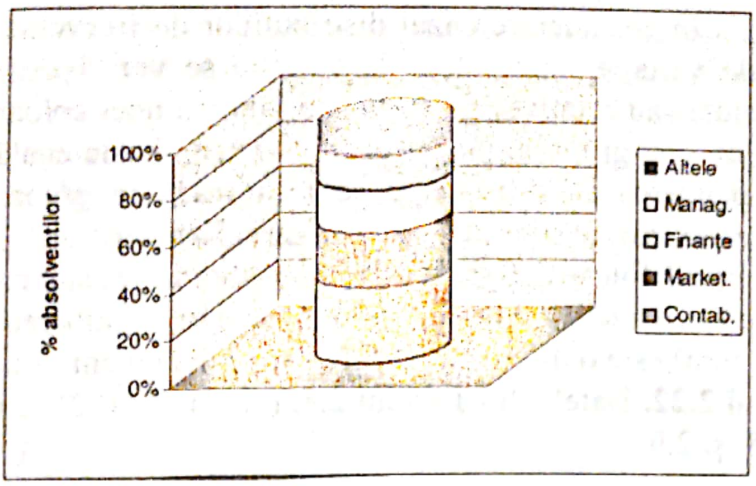


Figura 2.6. **Structura absolvenților** pe domenii de ocupare
(diagramă de structură **tridimensională** prin cilindru)

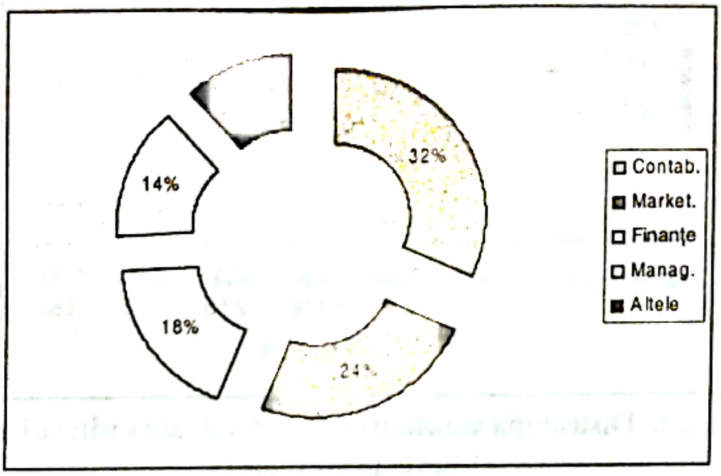


Figura 2.7. Structura absolvenților pe domenii de ocupare
(diagramă de structură bidimensională prin cerc „explodat")

D. Histograma, poligonul frecvențelor și curbele frecvențelor cumulate

Pentru reprezentarea grafică a distribuțiilor heterograde de frecvențe, pe intervale de variație se folosesc histograma, poligonul frecvențelor și curbele frecvențelor cumulate.

Histograma se trasează într-un sistem de axe rectangulare și apare sub forma unei succesiuni de coloane, lipite dacă intervalele sunt continue, de lățime egală sau diferită (după cum intervalele sunt egale sau nu).

Principala diferență între histogramă și graficul prin coloane constă în tipul variabilei măsurată pe axa Ox: dacă la graficul prin coloane pe această axă se reprezintă o variabilă calitativă, la histogramă se reprezintă o variabilă cantitativă continuă, măsurată pe scală de interval sau de raport. Histograma conține pe axa Oy frecvențe absolute sau relative.

Scanned with OKEN Scanner

Luăm întâi în considerare cazul distribuţiilor de frecvenţe heterograde pe intervale egale de variaţie. Intervalele de variaţie se vor reprezenta pe Ox, iar frecvenţele (absolute sau relative) pe Oy. Deci lăţimea unei coloane este egală cu mărimea intervalului de grupare, în timp ce înălţimea ei este egală cu numărul de unităţi (sau cu ponderea acestora) care se încadrează în grupa respectivă, din punctul de vedere al caracteristicii de sistematizare. Reprezentările grafice vor avea aceeaşi alură în cazul folosirii frecvenţelor absolute sau relative. Pe ambele axe este permisă întreruperea scării de reprezentare, dacă între valoarea minimă a variabilei şi originea (zero) este o diferenţă mare, comparativ cu amplitudinea variaţiei.

Exemplul 2.22. Datele din Tabelul 2.8. (coloanele 0-2) se reprezintă grafic ca în Figurile 2.8. şi 2.9.

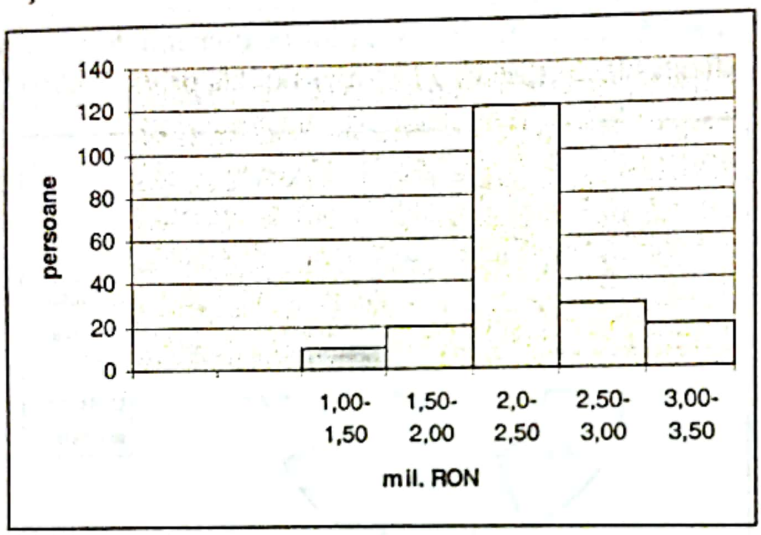


Figura 2.8. Distribuţia vânzătorilor după valoarea vânzărilor săptămânale (frecvenţe absolute)

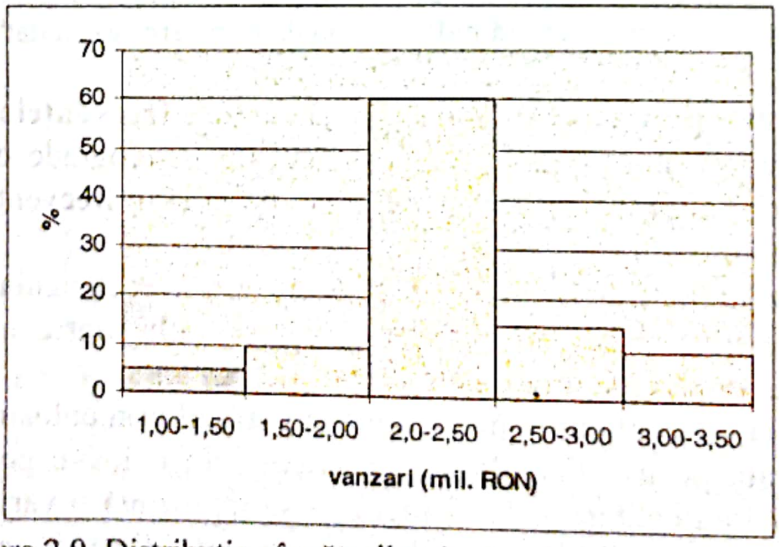


Figura 2.9. Distribuţia vânzătorilor după valoarea vânzărilor săptămânale (frecvenţe relative)

Scanned with OKEN Scanner

În cazul distribuţiilor de frecvenţe pe intervale neegale de variaţie, coloanele vor avea lăţimi diferite. Aria unei coloane ridicate pentru o anumită grupă este proporţională cu frecvenţa grupei respective, de aceea, dacă s-a mărit lăţimea coloanelor faţă de cazul grupării pe intervale egale (deoarece intervalele neegale s-au format prin alipirea a două, trei sau mai multe intervale egale), trebuie micşorate înălţimile coloanelor (adică în loc de frecvenţele absolute obişnuite, vom reprezenta pe Oy frecvenţele reduse la unitate).

Exemplul 2.23. Pentru datele din Tabelul 2.9. s-a realizat histograma din Figura 2.10.

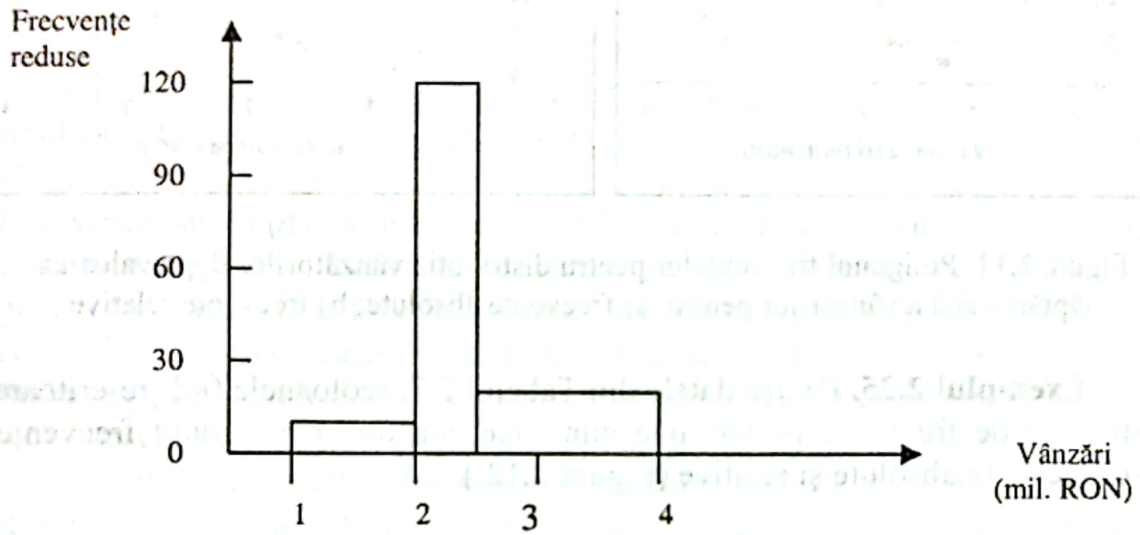
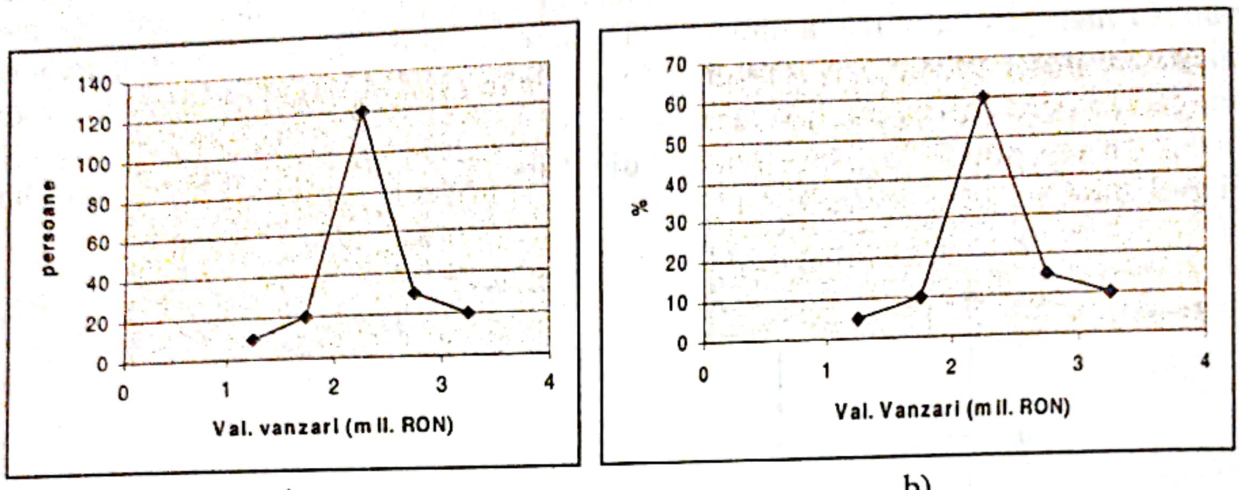


Figura 2.10. Distribuţia vânzătorilor după valoarea vânzărilor

Poligonul frecvenţelor este şi el utilizat pentru reprezentarea grafică a distribuţiilor heterograde de frecvenţe, atunci când sistematizarea datelor s-a făcut după o caracteristică numerică continuă sau discontinuă. Pentru construirea lui, din fiecare valoare a caracteristicii sau din fiecare centru de interval se ridică câte o perpendiculară şi se marchează pe ea punctul aflat la o distanţă egală cu frecvenţa variantei sau intervalului respectiv. Unind toate punctele astfel găsite, împreună cu centrele unor intervale imaginare (unul înaintea primului interval real de grupare, celălalt imediat după ultimul interval de grupare), situate chiar pe axa Ox (deoarece aceste intervale „imaginare" au frecvenţe nule), rezultă un poligon numit *poligonul frecvenţelor*. Poligonul frecvenţelor se poate suprapune peste histogramă în cadrul aceluiaşi grafic, sau se poate trasa într-un grafic separat. În primul caz, punctele poligonului se vor afla pe bazele superioare ale coloanelor histogramei. Aria totală de sub histogramă va fi egală cu aria de sub poligon. Poligonul frecvenţelor se poate trasa atât pentru o distribuţie de frecvenţe absolute, cât şi pentru o distribuţie de frecvenţe relative, imaginile celor două poligoane fiind asemănătoare.

Scanned with OKEN Scanner

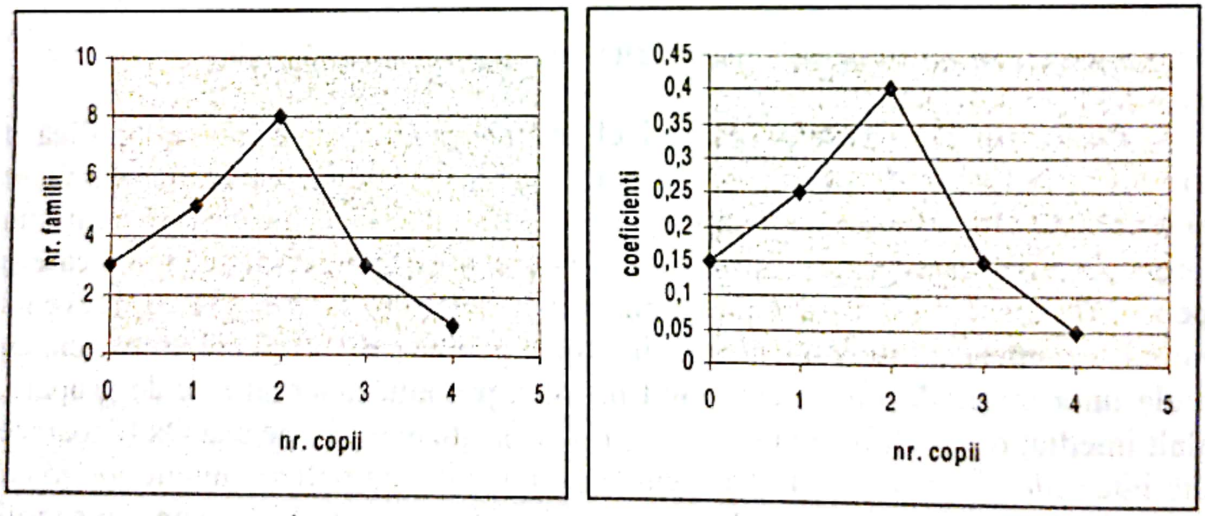
Exemplul 2.24. Pentru datele din Tabelul 2.8., ce prezintă o distribuție de frecvențe absolute și relative pe intervale egale de variație, s-a construit poligonul frecvențelor ca în Figura 2.11.



a) b)

Figura 2.11. Poligonul frecvențelor pentru distribuția vânzătorilor după valoarea săptămânală a vânzărilor pentru: a) frecvențe absolute; b) frecvențe relative

Exemplul 2.25. Pentru datele din Tabelul 2.7. (coloanele 0-2) referitoare la o distribuție de frecvențe pe variante numerice, s-a trasat poligonul frecvențelor pentru frecvențe absolute și relative (Figura 2.12.).



a) b)

Figura 2.12. Poligonul frecvențelor pentru distribuția familiilor după numărul de copii pentru: a) frecvențe absolute; b) frecvențe relative

Scanned with OKEN Scanner

Histograma şi poligonul frecvenţelor ne oferă informaţii privind forma unei distribuţii, gradul şi sensul asimetriei. Vom detalia analiza distribuţiilor pe baza reprezentărilor grafice în capitolul următor.

Curbele frecvenţelor cumulate (ogivele), numite şi *curbele cumulative ale frecvenţelor*, reprezintă o a treia modalitate de reprezentare grafică a distribuţiilor de frecvenţe pe intervale de variaţie sau pe variante. Ele se trasează atât pentru distribuţii de frecvenţe absolute, cât şi pentru distribuţii de frecvenţe relative.

Vom considera întâi cazul distribuţiilor de frecvenţe pe intervale de variaţie. Am văzut anterior că frecvenţele absolute/relative cumulate crescător se interpretează în relaţie cu limitele superioare ale intervalelor, şi că frecvenţele absolute/relative cumulate descrescător se interpretează în funcţie de limitele inferioare ale intervalelor de variaţie.

Pentru construirea curbei frecvenţelor cumulate crescător se marchează primul punct exact în prima limită inferioară, chiar pe axa Ox, deoarece în acel punct frecvenţa este nulă; în continuare, se marchează puncte în dreptul fiecărei limite superioare a intervalelor, până la înălţimea egală cu frecvenţa cumulată a intervalului respectiv. Unind toate aceste puncte se obţine o curbă crescătoare (ogivă) care are un punct de inflexiune. Aria de sub porţiunea de curbă cu panta mai înclinată desemnează intervalele în care au fost repartizate mai multe observaţii şi, dimpotrivă, ariile de sub porţiunile de curbă cu înclinaţie mai redusă indică intervalele cu frecvenţe mai mici, deci cu un număr mai mic de observaţii.

Asemănător curbei frecvenţelor cumulate crescător, se trasează şi curba frecvenţelor cumulate descrescător, pe acelaşi grafic sau pe grafice diferite. Diferenţa faţă de prima curbă este că acum se marchează punctele duse în dreptul limitelor inferioare ale intervalelor şi că se obţine o curbă descrescătoare (ogivă).

Exemplul 2.26. Pe baza datelor din Tabelul 2.8. (coloanele 0, 3, 4) s-au trasat curbele frecvenţelor absolute cumulate crescător şi descrescător (Figura 2.13.).

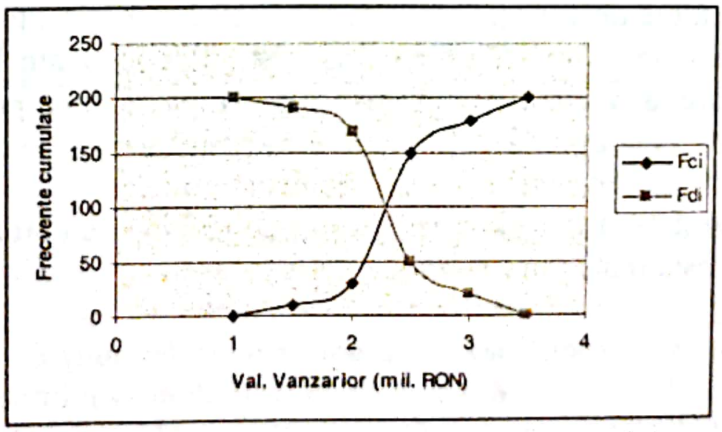


Figura 2.13. Curbele frecvenţelor cumulate crescător şi descrescător, pentru distribuţia vânzătorilor după valoarea vânzărilor săptămânale

Scanned with OKEN Scanner

În cazul unei distribuții de frecvențe după o variabilă numerică discretă, curba frecvențelor cumulate va avea aspectul unei scări, deoarece nici o unitate statistică nu va avea valori intermediare, situate între valorile discrete ale caracteristicii.

Exemplu 2.27. pentru distribuția familiilor după numărul de copii prezentată în Tabelul 2.7. (coloanele 0, 3, 4), s-a trasat curba frecvențelor cumulate crescător (Figura 2.14.).

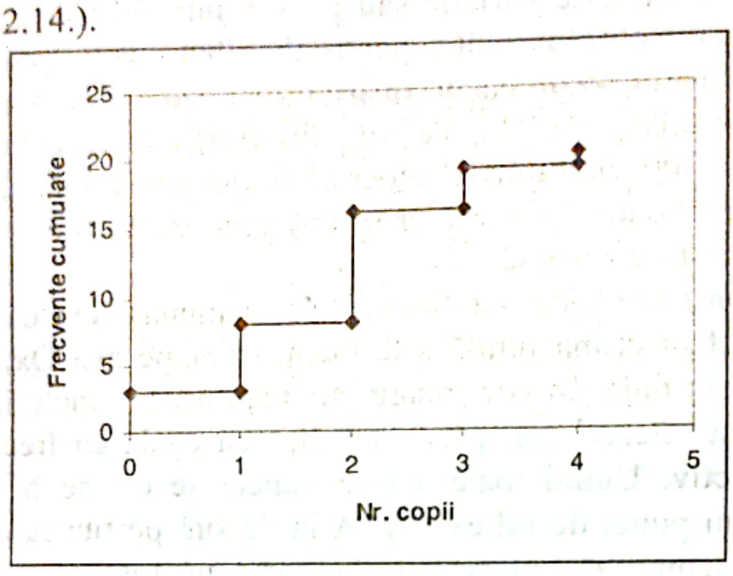


Figura 2.14. Frecvențele cumulate crescător pentru distribuția familiilor după numărul de copii

E. Diagrama norilor de puncte (graficul de corelație)

Diagrama norilor de puncte, numită şi corelogramă sau diagramă de împrăştiere reprezintă grafic relația dintre două sau mai multe variabile statistice, fiind deci utilizată în cazul distribuțiilor bi şi multidimensionale.

Considerând cazul unei distribuții bivariate, graficul se trasează într-un plan, în sistemul de coordonate rectangular Ox, Oy; pe axa Ox se reprezintă valorile sau intervalele de variație ale caracteristicii independente X, iar pe axa Oy valorile sau intervalele de variație ale caracteristicii dependente Y. Fiecare unitate statistică purtătoare a variabilelor X şi Y este reprezentată printr-un punct în cadranul format de cele două axe (de obicei cadranul întâi), de coordonate (x_i, y_i). Vom prezenta mai detaliat acest tip de grafic în capitolul .....

Exemplul 2.28. Reprezentarea grafică sub formă de corelogramă a datelor din Tabelul 2.14. este redată în Figura 2.15.

Scanned with OKEN Scanner

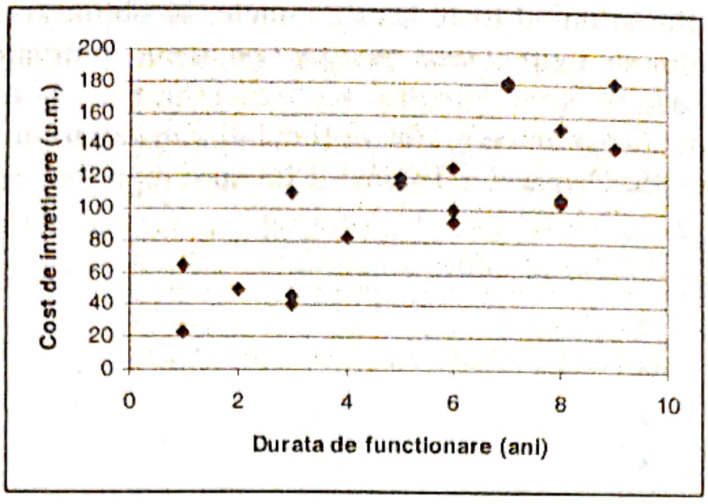


Figura 2.15. Corelogramă

Exemplul 2.29. Datele prezentate în Tabelul 2.15. sunt reprezentate grafic cu ajutorul corelogramei din Figura 2.16.

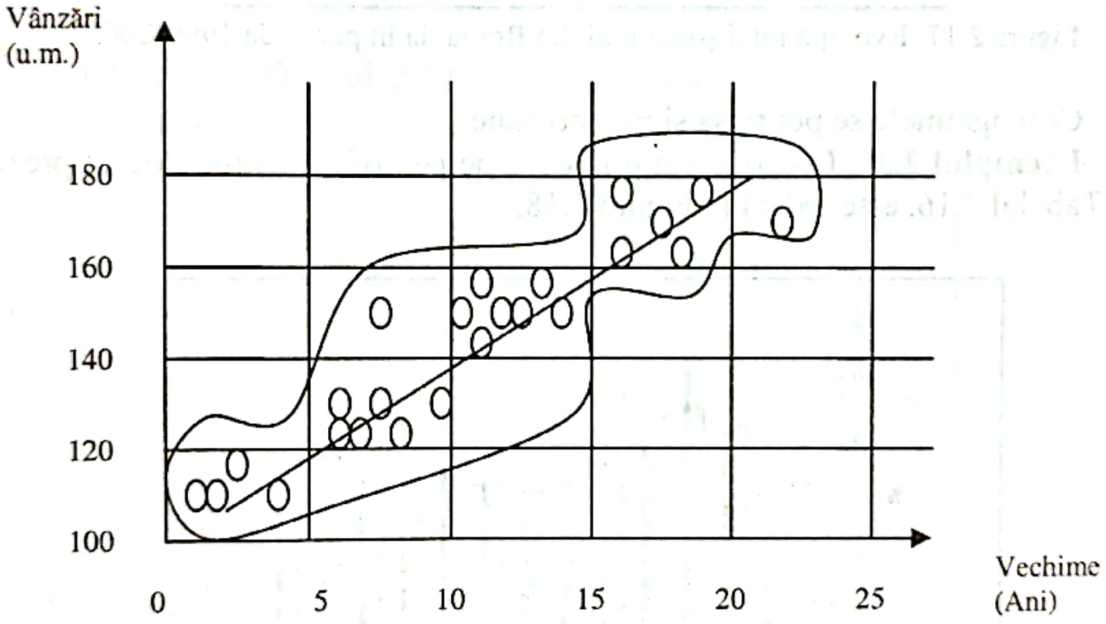


Figura 2.16. Corelogramă

F. Cronograma (historiograma)

Cronograma (numită și historiogramă) este tipul de reprezentare grafică utilizată în cazul seriilor cronologice. Ea se trasează în sistemul de axe rectangulare (Ox, Oy), marcându-se pe Ox momentele sau intervalele de timp, iar pe Oy valorile indicatorului a cărui evoluție este analizată. Dacă pe axa Ox avem momente de timp, atunci se ridică câte un punct din fiecare moment până la o înălțime egală cu

Scanned with OKEN Scanner

valoarea indicatorului şi unind toate aceste puncte, se obţine o curbă care reflectă evoluţia fenomenului în timp. Dacă pe axa Ox avem intervale de timp, atunci punctele se marchează în dreptul centrelor de interval, până la o înălţime egală cu nivelul fenomenului. Acest tip de grafic va fi detaliat în capitolul 6.

Exemplul 2.30. Datele din Tabelul 2.16. sunt reprezentate grafic cu ajutorul cronogramei din Figura 2.17.

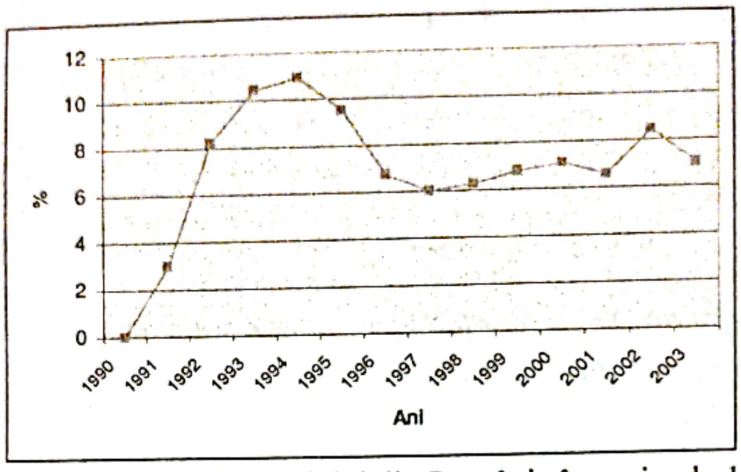


Figura 2.17. Evoluţia ratei şomajului din România în perioada 1990-2003

Cronogramele se pot trasa şi prin coloane.

Exemplul 2.31. Cronograma prin coloane pentru seria cronologică prezentată în Tabelul 2.16. este redată în Figura 2.18.

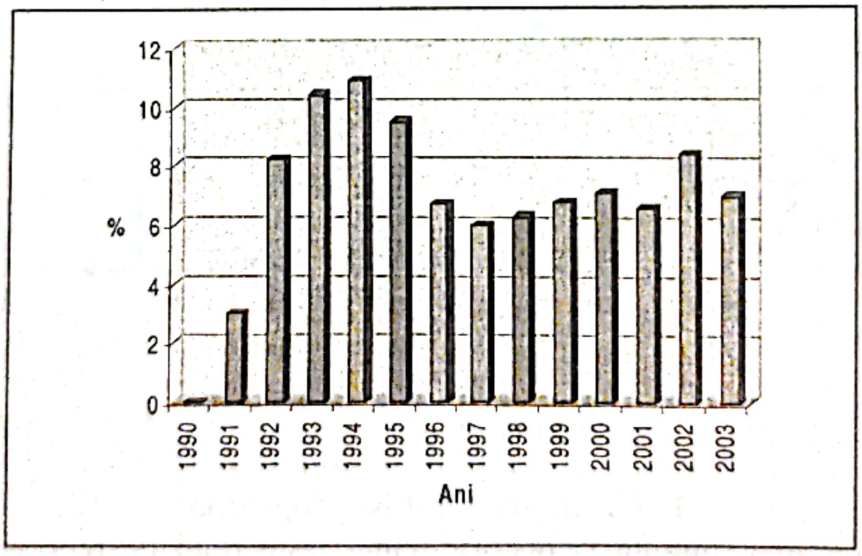


Figura 2.18. Evoluţia ratei şomajului din România în perioada 1990-2003

Scanned with OKEN Scanner

G. Diagrama polară (radială)

Diagrama polară (radială) se foloseşte pentru reprezentarea grafică a seriilor cronologice, atunci când evoluţia în timp a unui fenomen este şi rezultatul influenţei factorilor sezonieri. Acest tip de grafic se construieşte cu ajutorul reţelelor radiale; ciclurile periodice din cadrul perioadei analizate (zile, săptămâni, luni, trimestre etc.) sunt marcate pe circumferinţa cercului, în timp ce termenii seriei cronologice, care reflectă nivelul indicatorului urmărit, sunt marcaţi pe razele cercului.

Exemplul 2.32. Pentru o secţie a unei uzine se dau numărul defecţiunilor zilnice ale unei linii complexe de producţie (Tabelul 2.19.):

Tabelul 2.19.

Ziua	Număr de defecţiuni din:		
	Săptămâna 1	Săptămâna 2	Săptămâna 3
Luni	3	5	6
Marţi	4	4	4
Miercuri	7	7	5
Joi	8	10	8
Vineri	10	12	9

Datele din Tabelul 2.19. se reprezintă cu ajutorul unei diagrame radiale (polare) în Figura 2.19.:

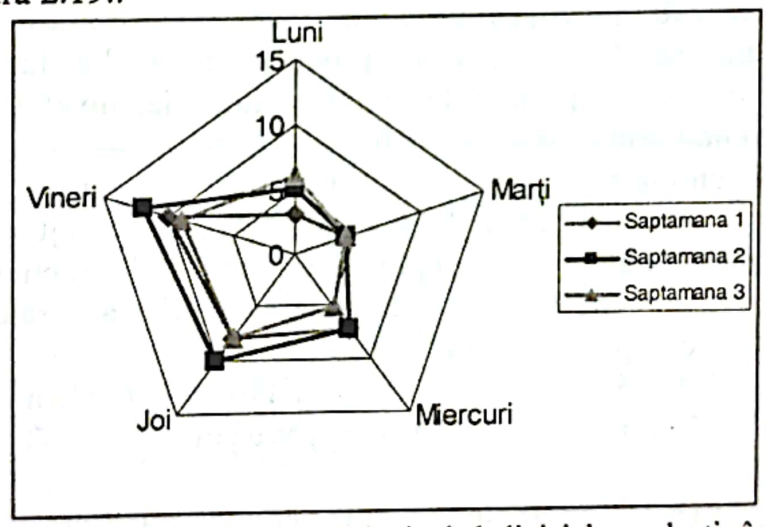


Figura 2.19. Evoluţia numărului zilnic de defecţiuni ale liniei de producţie în săptămânile 1-3

H. Diagrama prin suprafeţe

Diagramele prin suprafeţe se pot trasa cu ajutorul figurilor geometrice unidimensionale (care depind de o singură dimensiune de bază: cerc, pătrat etc.), bidimensionale (care sunt definite de câte două dimensiuni de bază, cum ar fi, dreptunghiul), sau al figurilor geometrice tridimensionale (paralelipiped dreptun-

Scanned with OKEN Scanner

ghic) etc. Denumirea graficului provine din faptul că suprafața figurii geometrice este proporțională cu valoarea indicatorului ce trebuie reprezentat.

Figurile geometrice unidimensionale se folosesc atunci când reprezentăm grafic fenomene independente, care nu sunt influențate de alți factori. În acest caz, pentru a trasa figura geometrică, se pune problema determinării dimensiunii definitorii a acesteia (a laturii pătratului, a razei cercului etc.).

Determinarea laturii pătratului se face parcurgând următorii pași:

- se extrage radicalul din mărimea indicatorului ce trebuie reprezentat grafic;
- se împarte rezultatul obținut la pasul anterior la unitatea de lungime a scării de reprezentare.

Determinarea razei cercului se face parcurgând următorii pași:

- se extrage radicalul din raportul între mărimea indicatorului ce trebuie reprezentat grafic și π (3,14);
- se împarte rezultatul obținut la pasul anterior la unitatea de lungime a scării de reprezentare.

Cu ajutorul diagramelor de suprafață se reprezintă grafic datele centralizate, totalizatoare, la nivel de colectivitate totală sau al grupelor acesteia.

Diagramele de suprafață trasate cu ajutorul figurilor geometrice bidimensionale (dreptunghiul) se folosesc atunci când studiem fenomene complexe (dependente), aflate sub influența unor factori independenți (2 factori). În acest caz, variabila complexă se descompune în produsul celor doi factori. Nivelurile factorilor se vor reprezenta pe laturile dreptunghiului, iar nivelul variabilei complexe va fi reprezentat prin suprafața acestuia.

În cazul fenomenelor complexe care se descompun în produsul a trei factori, se poate folosi diagrama de volum, elaborată cu ajutorul unor figuri geometrice tridimensionale (de exemplu, paralelipipedul dreptunghic). Cei trei factori se vor reprezenta pe lungimea, lățimea și înălțimea paralelipipedului, iar nivelul fenomenului complex prin volumul acestuia.

Exemplul 2.33. Se consideră valoarea Produsului Intern Brut pe locuitor (PIB/loc.) în anul 2002 (mii euro pe persoană), în patru țări europene (Tabelul 2.20.):

PIB/loc. în unele țări europene în anul 2002

Tabelul 2.20.

Țara	PIB/loc. (mii euro/pers.)
Franța	23,99
Marea Britanie	26,56
Luxemburg	55,00
Grecia	11,93

SURSA: EUROSTAT.

Scanned with OKEN Scanner

Pentru trasarea diagramei de suprafață cu ajutorul unor pătrate, vom determina conform algoritmului stabilit, laturile celor patru pătrate (se va trasa câte un pătrat pentru fiecare țară).

$$L_{Franta} = \frac{\sqrt{23,99}}{2} = 2,44 \text{ cm}; \qquad L_{Luxemb} = \frac{\sqrt{55}}{2} = 3,71 \text{cm}$$

$$L_{MareaBrit} = \frac{\sqrt{26,56}}{2} = 2,6 \text{ cm}; \qquad L_{Grecia} = \frac{\sqrt{11,93}}{2} = 1,73 \text{ cm}$$

Unitatea de lungime a scării de reprezentare a fost considerată 1 cm = 2 mii euro/pers. Graficul rezultat este trasat în Figura 2.20.

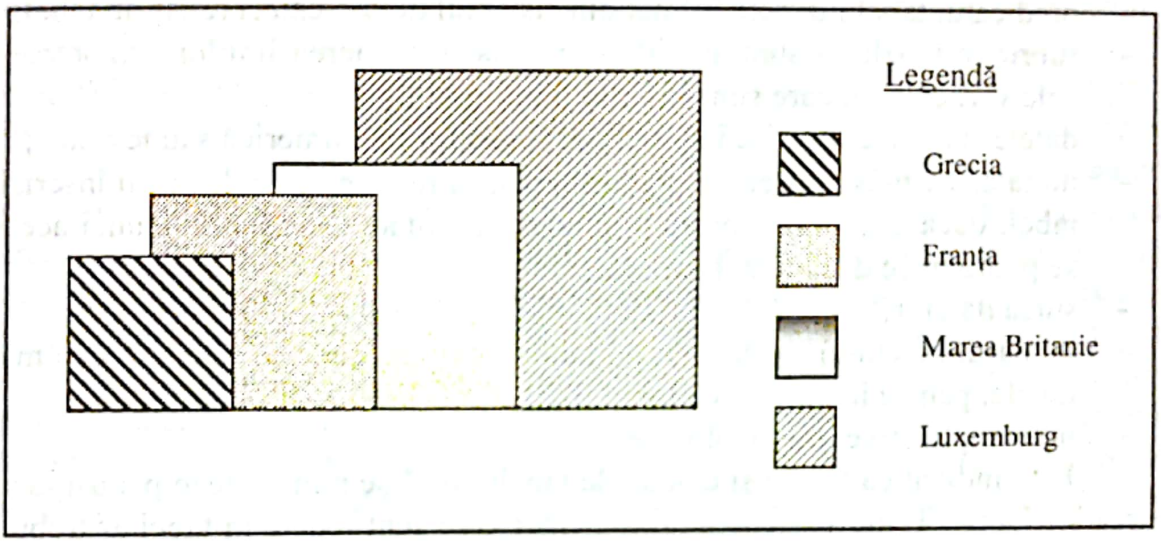


Figura 2.20. PIB/loc. în unele țări europene în anul 2002 (diagrama de suprafață)

În afară de tipurile de grafice prezentate, în analizele statistice se mai folosesc şi alte tipuri: *diagrama box-and-whisker*, *diagrama stem-and-leaf* şi altele. Un set de date nu se reprezintă grafic printr-un singur tip de grafic, ci întotdeauna există mai multe tipuri, unele mai potrivite, altele mai puțin potrivite, în funcție de posibilitatea lor de a reda mai multe sau mai puține informații esențiale despre setul de date.

Tabele statistice

Aşa cum am arătat, alături de grafice, şi tabelele statistice joacă un rol important în prezentarea datelor, căci ele pot releva anumite aspecte pe care graficele nu le pot pune în valoare. În unele situații, este mai importantă prezentarea valorilor numerice ale datelor, decât o vizualizare grafică a acestora. În

Scanned with OKEN Scanner

felul acesta, tabelele reprezintă un instrument complementar graficelor, de prezentare rapidă şi eficientă a datelor, dar şi de sistematizare a acestora.

Tabelul statistic cuprinde una sau mai multe serii statistice, ai căror termeni sunt înscrişi într-o reţea de linii şi coloane. Pentru ca un tabel statistic să fie corect elaborat şi să-şi atingă scopul, trebuie să conţină un set de elemente obligatorii sau opţionale şi să respecte anumite reguli:

- titlul tabelului este un element obligatoriu plasat înaintea tabelului, care descrie clar şi concis conţinutul datelor pe care le cuprinde;
- macheta tabelului este o reţea de linii ce alcătuiesc rubricile tabelului;
- subiectul tabelului este format din populaţia la care se referă datele înscrise în tabel;
- predicatul tabelului este format din sistemul de indicatori redaţi în tabel;
- rubricile tabelului sunt spaţiile create la întretăierea liniilor orizontale cu cele verticale, în care sunt înscrise datele;
- datele statistice înscrise în tabel pot fi sub formă numerică sau textuală;
- unitatea de măsură trebuie precizată pentru fiecare din indicatorii înscrişi în tabel; dacă toţi sunt exprimaţi în aceeaşi unitate de măsură, atunci aceasta se poate trece deasupra tabelului;
- sursa datelor;
- numărul tabelului – este necesar mai ales atunci când se folosesc mai multe tabele, pentru identificarea lor;
- note explicative, metodologice.

Este indicat ca liniile şi coloanele tabelului să se numeroteze pentru identificarea lor facilă. Toate valorile unui indicator care sunt redate în tabel ar trebui să aibă acelaşi număr de zecimale; uneori, se poate realiza şi o ordonare a datelor, de la valoarea cea mai semnificativă până la cea mai nesemnificativă.

Se pot include în tabel şi valorile medii ale liniilor şi coloanelor (atunci când acestea au sens), pentru a evidenţia tendinţa generală şi diferenţele dintre grupuri. De asemenea, se pot concentra unele date mai importante într-o coloană, pentru a fi văzute mai repede şi mai uşor.

În funcţie de importanţa datelor prezentate, de etapa în care au fost obţinute, de rolul lor în cadrul investigaţiei statistice, tabelele pot fi: descriptive (care cuprind, în general, date primare), de prelucrare, combinate, pe grupe (în care sunt prezentate datele obţinute în urma sistematizării şi centralizării lor) etc.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 3
ANALIZA DESCRIPTIVĂ A SERIILOR STATISTICE UNIVARIATE

3.1. Principalele aspecte ale analizei statistice a seriilor univariate

În capitolul anterior am văzut cum putem aplica diverse metode de organizare, de ordonare a unor largi seturi de date, într-o formă mai uşor de interpretat şi manipulat. (Aşa-numitele *metode exploratorii* care fac parte din statistica descriptivă.) Pe măsură ce avansăm în analiză, este necesar să aplicăm metode cantitative specifice, pentru a descrie şi sintetiza şi alte aspecte ale setului de date.

De exemplu, atunci când avem un număr mare de observaţii, identificăm întâi valorile minime şi maxime, apoi, prin gruparea/clasificarea datelor, formăm distribuţii de frecvenţe pe care le reprezentăm grafic. Pe baza graficului se caracterizează forma aproximativă a distribuţiei şi putem observa cam în ce zonă (în jurul cărei valori) se concentrează majoritatea observaţiilor. Desigur, toate informaţiile găsite până acum prin aplicarea metodelor de sistematizare, centralizare, prezentare şi reprezentare a datelor statistice prin serii, tabele şi grafice, sunt deosebit de utile, dar este necesară aplicarea, în continuare, a unor modele statistico-matematice pentru a calcula în mod exact şi obiectiv unele valori numerice sintetizatoare, ce descriu tendinţa centrală, care pot fi interpretate şi comparate unele cu altele. Acest lucru este deosebit de important, mai ales atunci când datele provin dintr-un eşantion, de la care trebuie să realizăm inferenţa la nivelul populaţiei totale.

Atunci când analizăm o distribuţie de frecvenţe, studiul se va concentra în trei direcţii:

– măsurarea tendinţei centrale;
– măsurarea variaţiei, a dispersării datelor;
– caracterizarea formei distribuţiilor de frecvenţe.

În felul acesta, un set foarte mare de date va putea fi caracterizat sub multiple aspecte doar cu ajutorul câtorva nivele numerice, pentru determinarea cărora vor fi aplicate modele matematice şi care vor constitui, la rândul lor, baza pentru analize statistice mult mai complexe.

3.2. Caracterizarea numerică a tendinţei centrale

Cele mai des întâlnite şi utilizate măsuri ale tendinţei centrale sunt: media, mediana şi modul. Într-un anume sens, toţi cei trei indicatori sunt sinonimi cu

termenul de „medie", deoarece ei încearcă să identifice, să localizeze tendința de mijloc a setului de date.

Toți cei trei indicatori sunt utilizați în mod uzual în analiza datelor statistice, folosirea lor depinzând, însă, de caracteristicile fiecăruia, de avantajele și dezavantajele sale, de tipul datelor analizate (dacă sunt date categoriale, ordinale sau cantitative; dacă sunt date discrete sau continue etc.) și de tipul și mărimea variabilității datelor (dacă ele sunt apropiate valoric unele de altele, sau există și unele nivele extreme, aberante).

- ***Media aritmetică*** este tipul de medie cel mai des întâlnit și utilizat și cel mai simplu de calculat. Ne referim adesea la cifra de afaceri medie, la numărul mediu de salariați, la rata medie a șomajului; uneori, însă, se creează ambiguitate, incertitudine cu privire la sensul termenului de „medie", în condițiile operării, de exemplu, cu date nominale (categoriale) sau ordinale, alteori, pe baza unor date numerice discrete se obțin valori ale mediei inexistente, practic, imposibil de atins (de exemplu, valori fracționare).

Media, în statistică, mai este definită drept nivelul la care ar fi ajuns un fenomen dacă toți factorii săi de influență ar fi acționat în mod constant.

Media aritmetică, din punct de vedere matematic, este valoarea care, înlocuind fiecare termen al unei serii, menține neschimbată suma acestora.

- ***Mediana*** reprezintă valoarea termenului din mijloc al unei serii ordonate de date. Indicatorul face parte din categoria indicatorilor de poziție ai unei serii, ca atare ea se determină numai pentru date între care se poate stabili o relație de ordine (date ordinale sau numerice, nu și pentru date nominale).
- ***Modul*** reprezintă valoarea care apare cel mai des într-o serie de date, adică valoarea cu frecvența maximă. Se poate determina pentru toate tipurile de date, indiferent de scala pe care se reprezintă ele, inclusiv pentru cele categoriale (calitative sau nominale).

Exemple de utilizare a modului: când vrem să determinăm care este atitudinea cea mai răspândită a unor persoane față de un proiect legislativ, față de un nou produs etc., care este cea mai căutată măsură de pantofi etc.

Modul este indicatorul cel mai ușor de determinat dintre toți indicatorii tendinței centrale, care se identifică, în general, prin observare, mai mult decât prin calcul statistic sau matematic.

3.2.1. Determinarea mediei, medianei și a modului pentru o serie de date negrupate

- ***Media aritmetică*** se determină plecând de la definiția dată anterior:

Fie o serie de date nesistematizate, de forma (x_i), $i = \overline{1,n}$. Suma termenilor seriei va fi:

Scanned with OKEN Scanner

$$S = x_1 + x_2 + ... + x_n = \sum_{i=1}^{n} x_i \qquad (3.1.)$$

Media aritmetică se notează, uzual, în statistică, cu $\bar{x}$. Dacă înlocuim fiecare termen al seriei cu media aritmetică, aceeaşi sumă se mai poate scrie:

$$S = \bar{x} + \bar{x} + ... + \bar{x} = n \cdot \bar{x} = \sum_{i=1}^{n} x_i \qquad (3.2.)$$

Rezultă de aici că:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (3.3.)$$

relaţie care descrie o medie aritmetică simplă, utilizată pentru o serie simplă, de date statistice negrupate. Ea se calculează ca raport între suma tuturor observaţiilor şi numărul acestora.

De cele mai multe ori, nu se cunosc valorile caracteristicii studiate pentru toate unităţile populaţiei, ci doar pentru o parte, pentru o submulţime a lor, deci pentru un eşantion. Vom nota deci, în cele ce urmează, media aritmetică a eşantionului cu $\bar{x}$.

Exemplul 3.1. numărul de piese realizat de cinci muncitori într-o zi este: 7, 8, 3, 9, 8. Numărul mediu de piese realizate de un muncitor în acea zi este:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{7+8+3+9+8}{5} = \frac{35}{5} = 7 \text{ piese/persoană.}$$

Dacă rezultatul ar fi fost o valoare fracţionară, atunci aceasta s-ar fi rotunjit la întregul cel mai apropiat (prin lipsă sau prin adaos), deoarece datele sunt discrete.

- ***Mediana***, pentru o serie de date negrupate, se determină parcurgând următorii paşi:

PAS 1: Se ordonează termenii seriei, crescător sau descrescător;

PAS 2: Se determină poziţia (rangul) termenului din mijlocul seriei, cu relaţia:

$$Poz\ Me = \frac{n+1}{2} \qquad (3.4.)$$

PAS 3: Se determină valoarea medianei, ca fiind egală cu termenul central de rang $\left[\frac{n+1}{2}\right]$. Pot apărea două situaţii:

Scanned with OKEN Scanner

a) Seria are număr impar de termeni.

În acest caz, în mijlocul seriei există un singur termen central, iar *Poz Me* este un număr întreg. Mediana va fi egală chiar cu termenul central.

Exemplul 3.2. Pentru seria din exemplul 3.1., ordonăm crescător termenii acesteia: 3, 7, 8, 8, 9.

$$Poz\,Me = \frac{n+1}{2} = \frac{5+1}{2} = 3\text{, deci } Me = x_{[3]} = 8$$

b) Seria are număr par de termeni.

În acest caz, seria va avea doi termeni centrali. Poziția medianei va fi un număr fracționar. Pot apărea două subcazuri:

b_1) Seria este compusă din date numerice.

Mediana este egală cu media aritmetică simplă a celor doi termeni centrali. Dacă datele sunt discrete şi valoarea calculată a medianei este fracționară, atunci se specifică cele două valori centrale, fără a mai calcula mediana ca valoare exactă.

Exemplul 3.3. Cifra de afaceri pentru 6 firme este (mil. RON): 25,3; 37,4; 30,5; 42,3; 35,6; 28,0.

Ordonăm crescător termenii seriei: 25,3; 28,0; 30,5; 35,6; 37,4; 42,3.

$$Poz\,Me = \frac{n+1}{2} = \frac{6+1}{2} = 3{,}5$$

Mediana se găseşte între al treilea şi al patrulea termen şi va fi egală cu:

$$Me = \frac{30{,}5+35{,}6}{2} = 33{,}05 \text{ mil. RON}$$

b_2) Seria este compusă din date calitative ordinale.

Mediana se găseşte între cei doi termeni centrali ai seriei, dar aceştia, fiind variante calitative, nu se poate calcula media lor aritmetică. Dacă cei doi termeni centrali sunt identici (au aceeaşi variantă), atunci mediana va fi egală chiar cu acea variantă. Dacă, dimpotrivă, cei doi termeni centrali conţin variante diferite, atunci se precizează ambii termeni ca valoare a medianei.

Exemplul 3.4. Calificativele obținute de 6 studenți la un proiect sunt: „satisfăcător", „bine", „insuficient", „foarte bine", „satisfăcător", „excelent".

Se ordonează crescător termenii seriei:
„insuficient", „satisfăcător", „satisfăcător", „bine", „foarte bine", „excelent".

$$Poz\,Me = \frac{n+1}{2} = \frac{6+1}{2} = 3{,}5$$

Cei doi termeni centrali sunt al treilea şi al patrulea, adică „satisfăcător" şi „bine". Întrucât ei sunt diferiți, se precizează ambele variante centrale:

$Me \in \{„satisfacator", „bine"\}$

Scanned with OKEN Scanner

• Pentru o serie simplă, ***modul*** se determină căutând valoarea care apare cel mai des (cu frecvență maximă).

Într-o serie de date pot exista: o singură valoare modală [serie unimodală – Figura 3.1. a)], două valori modale [serie bimodală – Figura 3.1. b)], trei sau mai multe valori modale [serie multimodală – Figura 3.1. c)].

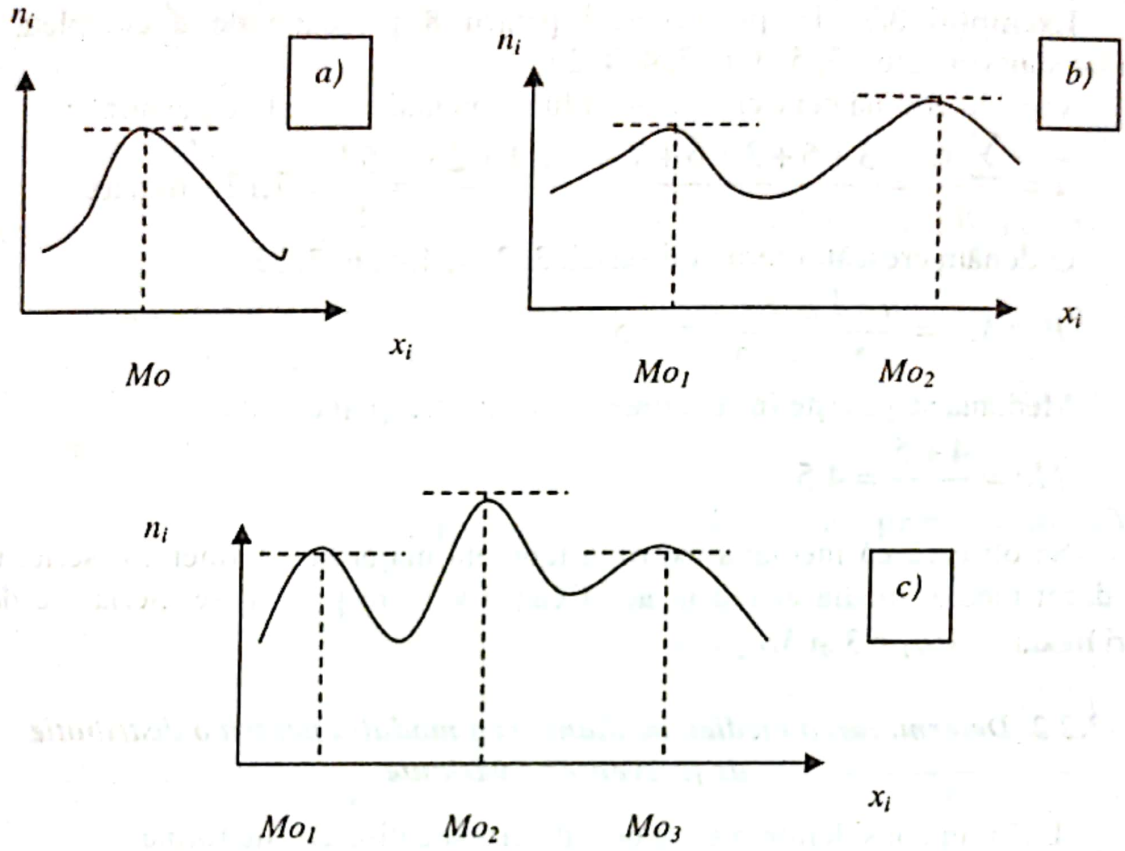


Figura 3.1. Serii cu număr diferit de valori modale:
a) unimodală; b) bimodală; c) multimodală

Exemplul 3.5. Într-un magazin de încălțăminte pentru femei, în decursul unei ore, 7 cliente au cumpărat pantofi cu următoarele mărimi: 36, 37, 35, 38, 37, 36, 37. Valoarea modală a acestei serii este 37, deoarece această variantă se repetă cel mai des (de 3 ori) şi arată mărimea de pantofi cea mai preferată de cliente.

Dacă într-o serie fiecare variantă apare o singură dată sau de acelaşi număr de ori ca şi celelalte variante, atunci determinarea modului este lipsită de sens.

Exemplul 3.6. 6 persoane au fost rugate să-şi spună părerea asupra unui produs alimentar nou promovat. Răspunsurile lor au fost: „deloc gustos", „gustos", „foarte gustos", „destul de gustos", „destul de gustos", „gustos".

Seria prezentată mai sus are două valori modale: „destul de gustos" şi „gustos", fiecare variantă apărând de două ori (frecvența maximă este 2); este o serie bimodală.

Scanned with OKEN Scanner

$Mo_1 = \{„destul\ de\ gustos“\}$

$Mo_2 = \{„gustos“\}$

Dacă într-o serie există şi o valoare extremă, atunci media aritmetică va fi influenţată de această valoare. Deoarece media este sensibilă faţă de aceste valori, este mai indicată caracterizarea tendinţei centrale a seriei cu ajutorul altui indicator, de exemplu al medianei, care este mai puţin afectată de valorile aberante.

Exemplul 3.7. Timpii necesari pentru 8 persoane de a completa un formular sunt (minute): 3, 5, 3, 6, 7, 4, 4, 25.

Vom determina cei trei indicatori fundamentali ai tendinţei centrale:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3+5+3+6+7+4+4+25}{8} = \frac{57}{8} = 7{,}125 \text{ minute}$$

Ordonăm crescător termenii seriei: 3, 3, 4, 4, 5, 6, 7, 25

$$Poz\ Me = \frac{n+1}{2} = \frac{8+1}{2} = 4{,}5$$

Mediana se găseşte între termenul al patrulea şi al cincilea.

$$Me = \frac{4+5}{2} = 4{,}5$$

Se observă că mediana exprimă tendinţa majorităţii termenilor seriei mai bine decât media, media având în acest caz o valoare prea mare. Seria are două valori modale: $Mo_1 = 3$ şi $Mo_2 = 4$.

3.2.2. Determinarea mediei, medianei şi a modului pentru o distribuţie de frecvenţe pe variante

Luăm în considerare o serie de date grupate discrete, de forma:

$$\begin{pmatrix} x_i \\ n_i \end{pmatrix}_{i=\overline{1,r}} = \begin{pmatrix} x_1 & x_2 \dots & x_i \dots & x_r \\ n_1 & n_2 \dots & n_i \dots & n_r \end{pmatrix} \qquad (3.5.)$$

unde: x_i este varianta discretă a caracteristicii X, înregistrată la unitatea statistică i;

n_i este frecvenţa absolută de apariţie a viariantei x_i înregistrată la unitatea statistică i.

- ***Media aritmetică*** se determină plecând de la aceeaşi definiţie ca cea prezentată anterior. Nivelul centralizat al variabilei X se poate scrie:

$$S = x_1 n_1 + x_2 n_2 + \dots + x_i n_i + x_r n_r = \sum_{i=1}^{r} x_i n_i \qquad (3.6.)$$

Scanned with OKEN Scanner

Înlocuind, conform definiției date, fiecare termen al seriei cu media, suma devine:

$$S = \bar{x} \cdot n_1 + \bar{x} \cdot n_2 + \ldots + \bar{x} \cdot n_i + \bar{x} \cdot n_r = \sum_{i=1}^{r} x_i n_i \qquad (3.7.)$$

de unde rezultă:

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} \qquad (3.8.)$$

relație care definește o medie aritmetică ponderată, utilizată în cazul distribuțiilor de frecvențe absolute.

Exemplul 3.8. Un eșantion de 50 de persoane a fost inclus într-un studiu statistic, fiecare persoană fiind întrebată asupra numărului de ore în care s-a uitat la TV în ziua precedentă. Răspunsurile, grupate, se prezintă ca în Tabelul 3.1. (coloanele 0, 1).

Distribuția persoanelor după numărul de vizionare a programului TV

Tabelul 3.1.

Număr de ore de vizionare (ore) (x_i)	Număr de persoane (n_i)	$x_i n_i$	Ponderea persoanelor (%) n_i^*	$x_i n_i^*$	*Fci*
0	1	2	3	4	5
0,5	7	3,5	14	7	7
1,0	10	10	20	20	17
2,0	18	36	36	72	35
3,0	12	36	24	72	47
4,0	3	12	6	24	50
Total	50	97,5	100	195	–

Pentru a determina numărul mediu de ore de vizionare a programului TV pe o pesoană, se aplică formula mediei aritmetice ponderate pentru o distribuție de frecvențe absolute, conform relației (3.8.), în care x_i sunt numărul de ore de vizionare a programului TV, iar n_i reprezintă numărul persoanelor care au vizionat programul TV timp de x_i ore. Pentru ușurarea calculelor, se determină produsul $x_i n_i$ în coloana 2 a Tabelului 3.1. Rezultă că media va fi:

$$\bar{x} = \frac{97,5}{50} = 1,95 \text{ ore/pers.}$$

Scanned with OKEN Scanner

Dacă în relația (3.8.) împărțim și numărătorul și numitorul la n, relația devine:

$$\bar{x} = \frac{\frac{1}{n}\sum_{i=1}^{r} x_i n_i}{\frac{1}{n}\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} x_i \frac{n_i}{n}}{\sum_{i=1}^{r} \frac{n_i}{n}} = \frac{\sum_{i=1}^{r} x_i n_i^*}{\sum_{i=1}^{r} n_i^*} \qquad (3.9.)$$

unde: $$n_i^* = \frac{n_i}{\sum_{i=1}^{r} n_i} = \frac{n_i}{n} \qquad (3.10.)$$

sunt frecvențele relative de apariție ale variantelor x_i.

- dacă frecvențele relative sunt exprimate în coeficienți, atunci:

$$\sum_{i=1}^{r} n_i^* = 1 \qquad (3.11.)$$

Relația (3.9.) devine:

$$\bar{x} = \sum_{i=1}^{r} x_i n_i^* \qquad (3.12.)$$

- dacă frecvențele relative sunt exprimate procentual, atunci:

$$\sum_{i=1}^{r} n_{i\%}^* = 100 \qquad (3.13.)$$

Relația (3.9.) devine:

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_{i\%}^*}{100} \qquad (3.14.)$$

Exemplul 3.9. Dacă s-ar fi cunoscut frecvențele relative (n_i^*) în loc de cele absolute (vezi coloana 3, Tabelul 3.1.), media aritmetică s-ar fi determinat cu relația:

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i^*}{\sum_{i=1}^{r} n_i^*} = \frac{\sum_{i=1}^{r} x_i n_{i\%}^*}{100} = \frac{195}{100} = 1{,}95 \text{ ore/pers.}$$

obținându-se aceeași valoare ca și în cazul precedent (calculele intermediare se găsesc în coloana 4, Tabelul 3.1.)

Scanned with OKEN Scanner

- ***Mediana*** se determină în cazul distribuţiilor de frecvenţe pe variante parcurgând următorii paşi:

PAS 1: Se cumulează crescător frecvenţele.

PAS 2: Se determină poziţia medianei (rangul termenului central) cu relaţia:

$$Poz\ Me = \frac{\sum_{i=1}^{r} n_i + 1}{2} = \frac{n+1}{2} \qquad (3.15.)$$

Observaţie: dacă în loc de frecvenţe absolute avem frecvenţe relative, atunci poziţia medianei va fi o valoare fixă, egală cu:

$$Poz\ Me = \frac{\sum_{i=1}^{r} n_{i\%}^{*} + 1}{2} = \frac{101}{2} = 50{,}5$$

PAS 3: Se determină valoarea mediană, egală cu prima variantă a cărei frecvenţă cumulată este mai mare decât poziţia medianei.

Exemplul 3.10. Pentru distribuţia de frecvenţe din Tabelul 3.1., se determină valoarea mediană. Pentru aceasta, calculăm frecvenţele cumulate crescător (Tabelul 3.1., coloana 5). Poziţia medianei în serie va fi:

$$Poz\ Me = \frac{\sum_{i=1}^{r} n_i + 1}{2} = \frac{51}{2} = 25{,}5$$

Prima variantă a cărei frecvenţă cumulată este mai mare decât poziţia medianei este varianta „2", deoarece $Fc_{(2)} = 35 > 25{,}5$. Aşadar, varianta mediană este Me = 2, ceea ce înseamnă că jumătate din persoane au vizionat programul TV mai puţin de 2 ore, şi jumătate mai mult de 2 ore.

Asemănător se determină mediana şi pentru o distribuţie de frecvenţe pe variante, pentru o caracteristică ordinală.

Exemplul 3.11. Frecvenţa cu care 250 de locuitori ai unui cartier din oraşul X îşi fac cumpărăturile de la un supermarket din zonă este (Tabelul 3.2., coloanele 0, 1):

Distribuţia locuitorilor după frecvenţa efectuării cumpărăturilor

Tabelul 3.2.

Frecvenţa	Număr de locuitori	Frecvenţe cumulate
0	1	2
Zilnic	32	32
Săptămânal	120	152
Bilunar	75	227
Lunar	18	245
Foarte rar sau deloc	5	250
Total	250	–

Scanned with OKEN Scanner

Pentru a determina mediana, se cumulează crescător frecvenţele (Tabelul 3.2., coloana 2) şi se găseşte poziţia medianei în serie:

$$Poz\ Me = \frac{\sum_{i=1}^{r} n_i + 1}{2} = \frac{251}{2} = 125{,}5$$

Mediana va fi egală cu prima variantă a cărei frecvenţă cumulată este superioară poziţiei medianei în serie, adică „săptămânal". Aceasta înseamnă că aproximativ jumătate din locuitorii incluşi în investigaţia statistică îşi fac cumpărăturile din acel supermarket mai des decât o dată pe săptămână şi cealaltă jumătate mai rar.

- ***Modul*** corespunde variantei cu frecvenţa maximă din cadrul distribuţiei.

Exemplul 3.12. Pentru distribuţia de frecvenţe din Tabelul 3.1., valoarea modală este 2, deoarece ea apare de cele mai multe ori (frecvenţa maximă este 18).

3.2.3. Determinarea mediei, medianei şi a modului pentru o distribuţie de frecvenţe pe intervale de variaţie

- ***Media***

În cazul unei distribuţii de frecvenţe pe intervale de grupare, nu mai cunoaştem valorile exacte ale caracteristicii, înregistrate la nivelul fiecărei unităţi statistice. În aceste condiţii, vom determina valori estimative ale mediei şi ale celorlalţi indicatori ai tendinţei centrale. Pentru aceasta, se consideră adevărată următoare ipoteză: aceea că valorile individuale ale caracteristicii în interiorul unui interval de grupare sunt uniform distribuite, repartizate. În condiţiile valabilităţii acestei ipoteze, putem reprezenta fiecare interval de variaţie prin centrul său. Aşadar, în relaţiile (3.8.) sau (3.9.), valorile x_i vor fi centrele de interval, iar valorile n_i, respectiv n_i^* – frecvenţele de apariţie (absolute sau relative).

Exemplul 3.13. Distribuţia a 200 de agenţi economici cu capital privat, care îşi desfăşoară activitatea în ramura textile-pielărie, după numărul de ani de existenţă, este (Tabelul 3.3., coloanele 0, 1):

Distribuţia agenţilor economici după numărul de ani de activitate

Tabelul 3.3.

Număr de ani de activitate (ani)	Număr de agenţi economici (n_i)	Centre de interval (x_i)	$x_i n_i$	*Fci*
.0	1	2	3	4
sub 4	17	2	34	17
4 – 8	29	6	174	46
8 – 12	80	10	800	126
12 – 16	37	14	518	163
16 – 20	27	18	486	190
peste 20	10	22	220	200
Total	200	–	2232	–

Scanned with OKEN Scanner

Pentru a calcula numărul mediu de ani de existență în domeniu ai unui agent economic studiat, trebuie să închidem intervalele extreme (primul şi ultimul), precizându-le limitele necunoscute: astfel, primul interval este 0–4, iar ultimul este 20–24. Se determinmă apoi centrele de interval (x_i), ca medie aritmetică simplă a limitelor fiecărui interval, intervalele fiind continue (vezi coloana 2, Tabelul 3.3.). Media aritmetică va fi:

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} = \frac{2232}{200} = 11{,}16 \text{ ani}$$

Calculele intermediare se regăsesc în Tabelul 3.3., coloana 3. Aşadar, un agent economic are o vechime medie în activitate de 11,16 ani.

- ***Mediana*** se determină parcurgând următorii paşi:

PAS 1: Se cumulează crescător frecvențele.

PAS 2: Se determină rangul termenului central, adică poziția medianei în serie, cu relația (3.23.). Pentru colectivități de volum mare se poate folosi în vederea determinării locului medianei relația:

$$Poz\ Me = \frac{n}{2} \qquad (3.16.)$$

PAS 3: Se identifică intervalul median (cel în care se plasează mediana), ca find primul interval a cărui frecvență cumulată este mai mare decât poziția medianei în serie, deci pentru care este îndeplinită condiția:

$$Fc_{Me} > \text{Poz Me} \qquad (3.17.)$$

PAS 4: Se calculează valoarea medianei cu ajutorul unei formule de interpolare pe intervalul median, conform relației:

$$Me = x_{inf(Me)} + h_{Me} \cdot \frac{Poz\ Me - Fc_{(Me-1)}}{n_{Me}} \qquad (3.18.)$$

unde:

$x_{\text{inf}(Me)}$ este limita inferioară a intervalului median;

h_{Me} este mărimea intervalului median;

$Poz\ Me$ este locul (poziția) medianei în serie;

n_{Me} este frecvența intervalului median;

$Fc_{(Me-1)}$ frecvența cumulată a intervalului ante-median.

Aplicarea relației (3.18.) este valabilă şi pentru determinarea medianei unei distribuții de frecvențe relative, precedată de aceiaşi paşi ca cei prezentați anterior.

Scanned with OKEN Scanner

Din punct de vedere geometric, mediana se determină cu ajutorul graficului astfel:

- se trasează curba frecvențelor cumulate crescător (unindu-se punctele ce au drept coordonate: pe abscisă – limitele superioare ale intervalelor, pe ordonată – frecvențele absolute sau relative cumulate ale grupelor, prin linii drepte). Trasarea curbei cumulative a frecvențelor ca o succesiune de segmente de dreaptă se explică prin presupunerea distribuției uniforme a frecvențelor în interiorul grupelor;
- se determină poziția medianei și se marchează pe axa ordonatelor (Oy);
- pentru acest punct, se determină valoarea frecvenței cumulate aflată pe curbă;
- coborând din acest punct al ogivei o perpendiculară, locul în care ea întâlnește axa Ox va fi mediana.

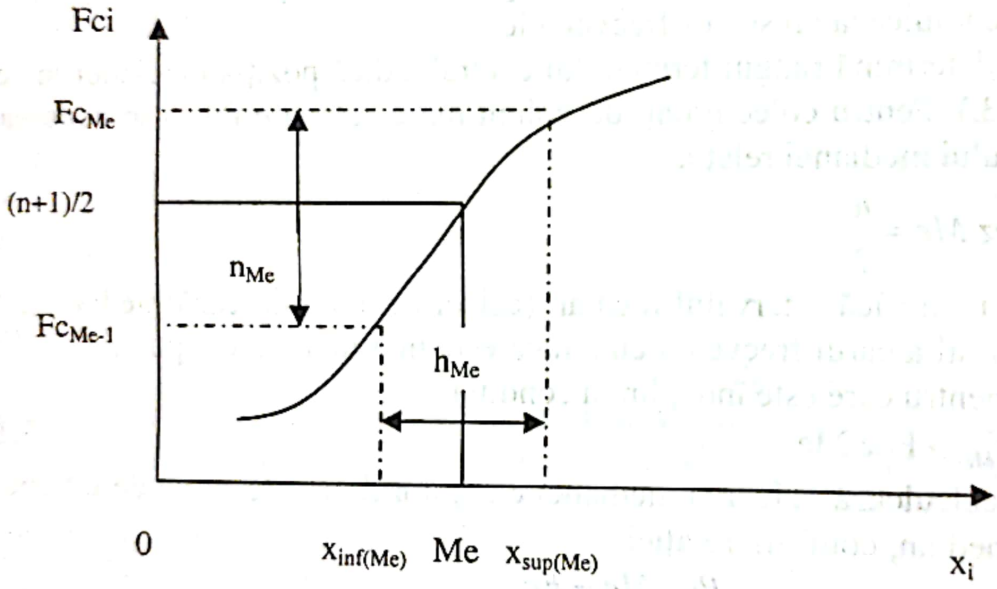


Figura 3.2. Determinarea pe cale grafică a medianei, pentru o distribuție de frecvențe pe intervale de valori

Exemplul 3.14. Pentru distribuția de frecvențe din Tabelul 3.3. (coloanele 0, 1) vrem să determinăm numărul median de ani de activitate al agenților economici incluși în studiu. Pentru aceasta s-au cumulat frecvențele absolute (Tabelul 3.3., coloana 4) și s-a determinat poziția medianei în cadrul seriei:

$$Poz\ Me = \frac{n+1}{2} = \frac{201}{2} = 100{,}5$$

Prima frecvență cumulată mai mare decât locul medianei este 126, deci mediana se găsește în intervalul 8-12. Pe acest interval aplicăm formula de interpolare:

Scanned with OKEN Scanner

$$Me = 8 + 4 \cdot \frac{100,5 - 46}{80} = 10,725 \text{ ani}$$

Aşadar, jumătate din firmele studiate activează în domeniu de mai puţin de 10,725 ani şi jumătate de peste 10,725 ani.

- ***Modul***

Atunci când setul de date este de dimensiuni reduse, modul nu este un indicator atât de semnificativ, întrucât completarea setului de date cu încă două sau trei valori poate schimba radical rezultatul. Nu la fel se întâmplă, însă, atunci când avem un număr mare de observaţii.

Exemplul 3.15. În Tabelul 3.4. este redată distribuţia salariaţilor după durata învoirilor din timpul programului de lucru:

Distribuţia angajaţilor după durata învoirilor

Tabelul 3.4.

Durata învoirilor (ore)	Număr de angajaţi (ni)
0	1
Sub 0,5	40
[0,5 – 1,0)	50
[1,0 – 1,5)	83
[1,5 – 2,0)	25
[2,0 – 2,5)	2
Total	200

Adăugarea câtorva valori nu va modifica valoarea modului, întrucât intervalul cu frecvenţa maximă este în mod clar (1,0-1,5) ore.

În cazul unei distribuţii de frecvenţe pe intervale de variaţie, modul se determină parcurgând următorii paşi:

PAS 1: se determină intervalul modal, ca fiind intervalul cu frecvenţa maximă;

PAS 2: asupra intervalului modal se aplică o formulă de interpolare, conform relaţiei:

$$Mo = x_{\inf(Mo)} + h_{Mo} \cdot \frac{d_1}{d_1 + d_2} \qquad (3.19.)$$

unde:

$x_{\inf(Mo)}$ este limita inferioară a intervalului modal;

h_{Mo} este mărimea intervalului modal;

d_1 este diferenţa dintre frecvenţa intervalului modal (n_{Mo}) şi frecvenţa intervalului anterior celui modal (n_{Mo-1}):

Scanned with OKEN Scanner

(3.20.)

$d_1 = n_{Mo} - n_{Mo-1}$

d_2 este diferența dintre frecvența intervalului modal (n_{Mo}) și frecvența intervalului următor celui modal (n_{Mo+1}):

(3.21.)

$d_2 = n_{Mo} - n_{Mo+1}$

În același mod, se determină modul și pentru o distribuție de frecvențe relative pe intervale de variație.

Pe grafic, modul se determină în felul următor (Figura 3.3.):

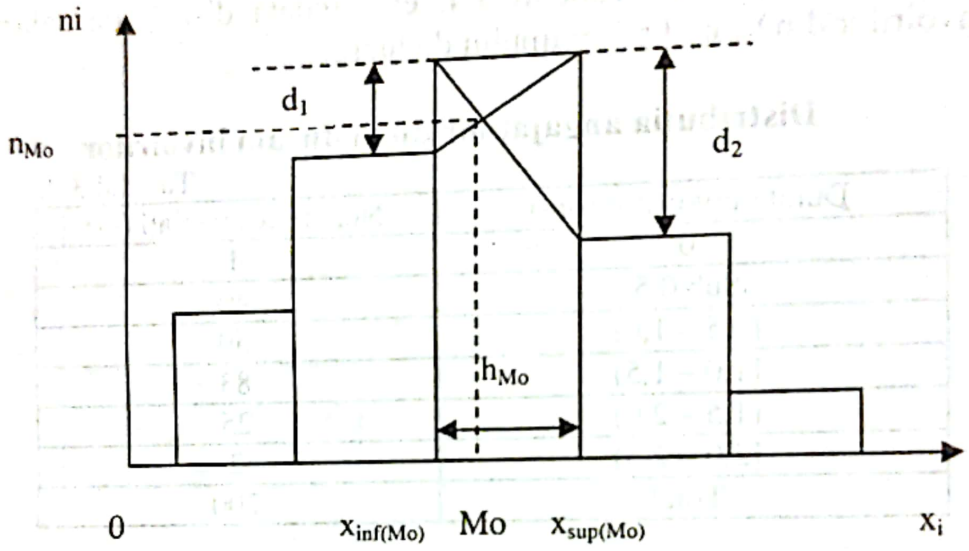


Figura 3.3. Determinarea pe cale grafică a modului, pentru o distribuție de frecvențe pe intervale de valori

Exemplul 3.16. Pentru distribuția din Tabelul 3.3., determinăm modul. Se observă că frecvența maximă este 80 (pentru al treilea interval) și deci indicatorul se regăsește în intervalul 8-12. Prin aplicarea relației (3.19.) obținem:

$$Mo = 8 + 4 \cdot \frac{51}{51+43} = 10{,}17 \quad \text{ani} \qquad (3.22.)$$

Așadar, cei mai mulți dintre agenții economici cu capital privat incluși în analiză au o vechime în activitate de aproximativ 10,17 ani.

3.2.4. Proprietățile mediei aritmetice

1) Media aritmetică este sensibilă la valorile extreme, aberante. De aceea, în ipoteza existenței unor asemenea valori, mult depărtate de ceilalți termeni ai seriei, media nu va mai reflecta corect tendința centrală a seriei, nu va mai fi reprezentativă. Variația valorilor observate va fi prea mare.

Scanned with OKEN Scanner

Exemplul 3.17. Valorile facturilor emise de o societate comercială într-o perioadă au fost: 902 RON, 1025 RON, 1100 RON, 1000 RON, 8230 RON, 950 RON. Valoarea medie a unei facturi emise în perioada analizată de către societatea comercială este:

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{902+1025+1100+1000+8230+9506}{6} = \frac{13207}{6} = 2201{,}17\,RON$$

Majoritatea termenilor seriei (mai precis 6 din 7) au valori cuprinse între 902 şi 1100 RON, dar media obţinută are o valoare superioară acestora, datorită existenţei unei valori extreme, mult depărtate de cele mai multe observaţii (8230 RON), care ridică valoarea mediei.

2) Media aritmetică se situează întotdeauna ca valoare între valorile extreme ale seriei pentru care s-a calculat:

$$x_{min} \le \bar{x} \le x_{max} \qquad (3.23.)$$

Exemplul 3.18. Pe baza datelor din exemplul 3.17., observăm că:

$$\left.\begin{array}{l} x_{min} = 902 \text{ RON;} \\ x_{max} = 8230 \text{ RON:} \\ \bar{x} = 2201{,}17 \text{ RON} \end{array}\right\} \Rightarrow x_{min} \le \bar{x} \le x_{max}$$

3) Media aritmetică calculată pentru o serie de valori egale este o valoare constantă, egală cu fiecare termen al seriei:

$$(\forall)\ (x_i = a), i = \overline{1,n},\ x_1 = x_2 = \ldots = x_n = \bar{x} = a \qquad (3.24.)$$

4) Suma tuturor abaterilor termenilor unei serii de la media lor aritmetică este întotdeauna zero. Altfel spus, abaterile aleatorii ale termenilor seriei faţă de media lor se compensează reciproc la nivelul ansamblului.

- Pentru o serie de date negrupate (simplă), proprietatea enunţată poate fi transpusă în termeni matematici astfel:

$$\text{Pentru } (\forall)\ (x_i), i = \overline{1,n}, \text{ avem } \sum_{i=1}^{n}\left(x_i - \bar{x}\right) = 0 \qquad (3.25.)$$

Într-adevăr, dacă desfacem parantezele în relaţia (3.25.), obţinem:

$$\sum_{i=1}^{n} x_i - n\bar{x} = 0. \qquad (3.26.)$$

Scanned with OKEN Scanner

$$\text{Cum } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \sum_{i=1}^{n} x_i = n \cdot \bar{x}. \qquad (3.27.)$$

Din relațiile (3.26.) și (3.27.) rezultă că: $n \cdot \bar{x} - n \cdot \bar{x} = 0$, relație adevărată.

- Pentru o serie de distribuție de frecvențe, proprietatea enunțată se transcrie matematic astfel:

Pentru o distribuție de frecvențe de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i = \overline{1,r}$, avem:

$$\sum_{i=1}^{r} (x_i - \bar{x}) n_i = 0 \qquad (3.28.)$$

Într-adevăr, prelucrând relația (3.28.), obținem:

$$\sum_{i=1}^{r} (x_i n_i - \bar{x} \cdot n_i) = \sum_{i=1}^{r} x_i n_i - \bar{x} \cdot \sum_{i=1}^{r} n_i = 0. \qquad (3.29.)$$

$$\text{Cum } \bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} \Rightarrow \sum_{i=1}^{r} x_i n_i = \bar{x} \cdot \sum_{i=1}^{r} n_i \,. \qquad (3.30.)$$

Din relațiile (3.29.) și (3.30.) reiese că $\bar{x} \cdot \sum_{i=1}^{r} n_i - \bar{x} \cdot \sum_{i=1}^{r} n_i = 0$. (3.31.)

Această proprietate este perfect adevărată numai dacă media rezultată din calcul este un cât exact.

Exemplul 3.19. Fie seria simplă $(x_i), i = \overline{1,n} = \{1,6,9,3,7,6\}$. Media aritmetică a seriei va fi:

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{32}{6} = 5,(3) \approx 5,33$$

Verificarea proprietății (4) a mediei aritmetice se face astfel:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = (1 - 5,33) + (6 - 5,33) + (9 - 5,33) + (3 - 5,33) + (7 - 5,33) +$$

$$+ (6 - 5,33) = -4,33 + 0,67 + 3,67 - 2,33 + 1,67 + 0,67 = 0,02 \approx 0$$

Scanned with OKEN Scanner

5) Dacă într-o serie (simplă sau de frecvenţe), pentru care s-a calculat media $\bar{x}$, fiecare termen se măreşte sau se micşorează cu o aceeaşi valoare c, atunci media noii serii, astfel obţinute, va fi şi ea modificată în acelaşi sens cu aceeaşi valoare c, faţă de media seriei iniţiale.

- Pentru o serie simplă, de forma $(x_i), i=\overline{1,n}$, pentru care s-a determinat media aritmetică cu relaţia (3.3.), modificăm termenii seriei iniţiale în sensul măririi lor, de exemplu, cu o constantă c. Atunci noua serie va fi de forma:

$$(y_i), i=\overline{1,n}, \text{ unde } y_i = x_i + c, (\forall)\ i=\overline{1,n} \qquad (3.32.)$$

Media noii serii va fi:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n}(x_i + c)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{n \cdot c}{n} = \bar{x} + c \qquad (3.33.)$$

- Pentru o serie de distribuţie de frecvenţe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i=\overline{1,r}$ pentru care s-a calculat media aritmetică cu relaţia (3.8.), dacă mărim fiecare termen al seriei cu o aceeaşi valoare c, se obţine o nouă serie de forma:

$$\begin{pmatrix} y_i \\ n_i \end{pmatrix}_{i=\overline{1,r}} = \begin{pmatrix} x_i + c \\ n_i \end{pmatrix}_{i=\overline{1,r}} \qquad (3.34.)$$

pentru care media aritmetică va fi:

$$\bar{y} = \frac{\sum_{i=1}^{r} y_i n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r}(x_i + c)n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} + \frac{c \cdot \sum_{i=1}^{r} n_i}{\sum_{i=1}^{r} n_i} = \bar{x} + c \qquad (3.35.)$$

La fel se demonstrează proprietatea (5) şi în cazul micşorării termenilor seriei cu o constantă.

Exemplul 3.20. Producţia zilnică realizată de 7 muncitori ai unui agent economic industrial a fost (bucăţi): 100, 85, 120, 118, 102, 90, 150. Producţia medie realizată de un muncitor este:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{765}{7} = 109{,}3 \approx 109 \text{ bucăţi}$$

Scanned with OKEN Scanner

În perioada următoare, fiecare muncitor a realizat zilnic cu 25 de piese mai mult. Noua producție medie zilnică a fost:

$$\bar{x}' = \frac{\sum_{i=1}^{n} x_i'}{n} = \frac{\sum_{i=1}^{n} (x_i + 25)}{n} = \bar{x} + 25 = 109{,}3 + 25 = 134{,}3 \approx 134 \text{ bucăți}$$

6) Dacă într-o serie (simplă sau de frecvențe) pentru care s-a calculat media $\bar{x}$, fiecare termen se măreşte sau se micşorează de un acelaşi număr de ori m, atunci media noii serii va fi şi ea modificată în acelaşi sens şi de acelaşi număr de ori m, faţă de media seriei inițiale.

- Pentru o sere simplă, de forma $(x_i), i = \overline{1,n}$, pentru care s-a determinat media aritmetică cu relația (3.3.), modificăm termenii seriei inițiale în sensul măririi lor, de exemplu, de m ori. Se obține, astfel, o nouă serie de forma:

$(y_i), i = \overline{1,n}$, unde $y_i = x_i \cdot m$, $(\forall)\ i = \overline{1,n}$ (3.36.)

Media noii serii va fi:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (x_i \cdot m)}{n} = \frac{m \cdot \sum_{i=1}^{n} x_i}{n} = m \cdot \bar{x} \qquad (3.37.)$$

- Pentru o serie de distribuție de frecvențe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i = \overline{1,r}$ pentru care s-a calculat media aritmetică cu relația (3.8.), dacă mărim fiecare termen al seriei de m ori, se obține o nouă serie de forma:

$$\begin{pmatrix} y_i \\ n_i \end{pmatrix}_{i=\overline{1,r}} = \begin{pmatrix} x_i \cdot m \\ n_i \end{pmatrix}_{i=\overline{1,r}} \qquad (3.38.)$$

pentru care media aritmetică va fi:

$$\bar{y} = \frac{\sum_{i=1}^{r} y_i n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} (m \cdot x_i) n_i}{\sum_{i=1}^{r} n_i} = \frac{m \cdot \sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} = m \cdot \bar{x} \qquad (3.39.)$$

La fel se demonstrează proprietatea (6) şi în cazul micşorării termenilor seriei de un anumit număr de ori.

Exemplul 3.21. Pe baza datelor din exemplul 3.20., să considerăm un nou compartiment pus în funcțiune al aceluiaşi agent economic, care are tot 7 muncitori. Fiecare dintre aceştia realizează zilnic doar 75% din producția zilnică a

Scanned with OKEN Scanner

muncitorilor din primul compartiment. Atunci producția medie a unui muncitor din noul compartiment este:

$$\bar{x}' = \frac{\sum_{i=1}^{n} x_i'}{n} = \frac{\sum_{i=1}^{n}\left(x_i \cdot \frac{75}{100}\right)}{n} = \frac{0{,}75 \cdot \sum_{i=1}^{n} x_i}{n} = 0{,}75 \cdot \bar{x} = 109{,}3 \cdot 0{,}75 = 81{,}975 \approx 82 \text{ bucăți}$$

Proprietățile (5) și (6) își dovedesc utilitatea atunci când se impune determinarea mediei unor serii de valori mari; în acest caz, prin aplicarea acestor proprietăți se vor obține serii „derivate" din primele, dar cu termeni de valori mai reduse, facilitându-se astfel calculele necesare determinării mediei.

7) Media aritmetică se mai poate determina și cu ajutorul așa-numitei „formule de calcul simplificat al mediei aritmetice":

- pentru seriile simple:

$$\bar{x} = \frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{h}\right)}{n} \cdot h + a \qquad (3.40.)$$

- pentru seriile de frecvențe:

$$\bar{x} = \frac{\sum_{i=1}^{r}\left(\frac{x_i - a}{h}\right) n_i}{\sum_{i=1}^{r} n_i} \cdot h + a \qquad (3.41.)$$

unde a și h sunt valori constante alese convenabil.

Relația (3.41.) se poate scrie asemănător și pentru frecvențele relative n_i^*:

$$\bar{x} = \frac{\sum_{i=1}^{r}\left(\frac{x_i - a}{h}\right) n_i^*}{\sum_{i=1}^{r} n_i^*} \cdot h + a \qquad (3.42.)$$

Pentru a demonstra validitatea relațiilor (3.40.) și (3.41.) vom prelucra expresiile din partea dreaptă a egalităților:

- pentru seriile simple:

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{h}\right)}{n} \cdot h + a = \frac{\frac{1}{h}\sum_{i=1}^{n}(x_i - a)}{n} \cdot h + a = \frac{1}{h}\left(\frac{\sum_{i=1}^{n} x_i}{n} - \frac{n \cdot a}{n}\right) \cdot h + a = \qquad (3.43.)$$

$$= \frac{1}{h}\left(\bar{x} - a\right) \cdot h + a = \bar{x}$$

Scanned with OKEN Scanner

- pentru seriile de frecvențe:

$$\frac{\sum_{i=1}^{r}\left(\frac{x_i-a}{h}\right)n_i}{\sum_{i=1}^{r}n_i}\cdot h+a=\frac{\frac{1}{h}\sum_{i=1}^{n}(x_i-a)n_i}{\sum_{i=1}^{r}n_i}\cdot h+a=\frac{1}{h}\left(\frac{\sum_{i=1}^{n}x_in_i}{\sum_{i=1}^{r}n_i}-\frac{a\cdot\sum_{i=1}^{r}n_i}{\sum_{i=1}^{r}n_i}\right)\cdot h+a=$$

$$=\bar{x}-a+a=\bar{x} \qquad (3.44.)$$

În cazul unei distribuții de frecvențe pe intervale egale de valori, pentru simplificarea calculelor, se alege drept a centrul intervalului cu frecvența maximă sau centrul intervalului median, iar drept h – mărimea intervalului.

Exemplul 3.22. Pentru distribuția din Tabelul 3.3. vom recalcula media folosind formula de calcul simplificat. Calculele intermediare se regăsesc în Tabelul 3.5.

Distribuția agenților economici după numărul de ani de activitate

Tabelul 3.5.

Număr de ani de activitate (ani)	Număr de agenți economici (n_i)	Centre de interval (x_i)	$x_i - a$	$\frac{x_i-a}{h}$	$\left(\frac{x_i-a}{h}\right)n_i$
0	1	2	3	4	5
sub 4	17	2	–8	–2	–34
4–8	29	6	–4	–1	–29
8–12	80	10	0	0	0
12–16	37	14	4	1	37
16–20	27	18	8	2	54
peste 20	10	22	12	3	30
Total	200	–	–	–	58

$a = 10$; $h = 4$.

Aplicând relația (3.41.), se determină media:

$$\bar{x}=\frac{58}{200}\cdot 4+10=11{,}16 \text{ ani}$$

Am obținut, așadar, aceeași valoare ca și în cazul utilizării formulei clasice a mediei.

8) Dacă toate frecvențele de apariție ale variantelor unei distribuții de frecvențe se simplifică sau se multiplică de k ori, atunci media seriei nu se modifică.

Scanned with OKEN Scanner

Fie o serie de distribuție de frecvențe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i = \overline{1,r}$ pentru care s-a calculat media aritmetică cu relația (3.8.). Prin simplificarea frecvențelor inițiale de k ori, se obține o nouă serie de forma:

$$\begin{pmatrix} x_i \\ n_i' \end{pmatrix}_{i=\overline{1,r}} = \begin{pmatrix} x_i \\ \frac{n_i}{k} \end{pmatrix}_{i=\overline{1,r}} \tag{3.45.}$$

Pentru noua distribuție, media aritmetică va fi:

$$\bar{x} = \frac{\sum x_i n_i'}{\sum n_i'} = \frac{\sum x_i \frac{n_i}{k}}{\sum \frac{n_i}{k}} = \frac{\frac{1}{k}\sum x_i n_i}{\frac{1}{k}\sum n_i} = \bar{x} \tag{3.46.}$$

Astfel se explică de ce **valorile mediei** aritmetice determinate pe baza frecvențelor absolute sau relative **sunt identice. Într**-adevăr, deoarece

$$n_i^* = \frac{n_i}{\sum n_i} = \frac{n_i}{n} \tag{3.47.}$$

înlocuirea frecvențelor absolute cu frecvențele relative în calculul mediei este echivalentă cu micşorarea tuturor frecvențelor absolute de n ori. În ambele cazuri, se obține aceeaşi valoare a mediei aritmetice.

Exemplul 3.23. Vârsta (în ani împliniți) a participanților la un curs postuniversitar, este (Tabelul 3.6. coloanele 0, 1):

Distribuția participanților la curs după vârstă

Tabelul 3.6.

Vârsta (ani)	Număr de participanți (n_i)	$x_i n_i$	n_i'	$x_i n_i'$
0	1	2	3	4
25	5	125	15	375
27	7	189	21	567
30	8	240	24	720
31	2	62	6	186
35	2	70	6	210
36	1	36	3	108
Total	25	722	75	2166

Vârsta medie a unui participant la curs va fi:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{722}{25} = 28{,}88 \approx 29 \text{ ani}$$

Scanned with OKEN Scanner

Dacă numărul participanţilor la curs, din fiecare categorie de vârstă se măreşte de 3 ori, atunci vârsta medie va deveni:

$$\bar{x} = \frac{\sum x_i n_i'}{\sum n_i'} = \frac{2166}{75} = 28{,}88 \approx 29 \text{ ani}$$

$$n_i' = 3 \cdot n_i$$

Rezultă că vârsta medie nu se modifică.

9) Suma pătratelor abaterilor termenilor seriei faţă de medie este mai mică decât suma pătratelor abaterilor termenilor seriei faţă de oricare altă valoare (reală sau ipotetică) a caracteristicii.

$$\min \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad (\forall) a \qquad (3.48.)$$

Exemplul 3.24. Numărul de personal pentru 5 agenţi economici la sfârşitul anului 2004 este (Tabelul 3.7. coloanele 0, 1):

Distribuţia agenţilor economici după numărul de personal

Tabelul 3.7.

Agentul economic	Număr de personal	$(x_i - 15)^2$	$(x_i - 23)^2$	$(x_i - 10)^2$	$(x_i - 32)^2$	$(x_i - \bar{x})^2$
0	1	2	3	4	5	6
A	20	25	9	100	144	0
B	15	0	64	25	289	25
C	23	64	0	169	81	9
D	10	25	169	0	484	100
E	32	289	81	484	0	144
Total	100	403	323	778	998	278

Numărul mediu de personal al unui agent economic este:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{100}{5} = 20 \text{ persoane}$$

În coloanele 2-5 din Tabelul 3.7. sunt calculate pătratele diferenţelor între termenii seriei, iar în coloana 6 a aceluiaşi tabel s-au determinat pătratele distanţelor între fiecare termen al seriei şi media aritmetică. Comparând totalurile coloanelor 2-6 din ultima linie a tabelului, se observă că:

$$\sum_{i=1}^{5} (x_i - \bar{x})^2 < \sum_{i=1}^{5} (x_i - 23)^2 < \sum_{i=1}^{5} (x_i - 15)^2 < \sum_{i=1}^{5} (x_i - 10)^2 < \sum_{i=1}^{5} (x_i - 32)^2$$

Scanned with OKEN Scanner

Alte tipuri de medii aritmetice

1. Un caz particular înregistrăm atunci când variabilele observate sunt calitative dihotomice. Colectivitatea este divizată, conform celor două variante calitative ale variabilei urmărite, în două grupe/clase de tipul: adevărat/fals, da/nu, bărbaţi/femei, căsătorit/necăsătorit, profit/pierderi etc. Deşi variantele unei asemenea caracteristici sunt de natură calitativă, se poate determina o valoare medie a acestora. Una din variante va fi considerată „stare favorabilă", în sensul că ea prezintă interes pentru cercetător. Această variantă va fi „codificată" numeric cu „1", iar cealaltă variantă („starea nefavorabilă") – cu „0". Pentru calculul mediei aritmetice ale unei asemenea variabile, datele pot fi sistematizate ca în Tabelul 3.8.

Sistematizarea datelor după o variabilă alternativă

Tabelul 3.8.

Variantele caracteristicii alternative	Codificare numerică (x_i)	Număr de unităţi statistice (frecvenţe) (n_i)	Frecvenţe relative (n_i^*)
DA (stare favorabilă)	1	m	$\frac{m}{n} = p$
NU (stare nefavorabilă)	0	$n-m$	$\frac{n-m}{n} = 1-p$
TOTAL	–	n	1

Media se va calcula conform relaţiei (3.15.), astfel:

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} = \frac{1 \cdot m + 0 \cdot (n-m)}{n} = \frac{m}{n} = p \qquad (3.49.)$$

Rezultă, aşadar, din relaţia (3.49.), că media aritmetică a unei variabile alternative este egală cu frecvenţa relativă în cazul favorabil, adică cu ponderea unităţilor statistice la care întâlnim varianta „favorabilă" a caracteristicii urmărite în volumul total al populaţiei.

Exemplul 3.25. În urma intervievării a 300 de vizitatori ai unui târg de turism, 75 de persoane au declarat că au intenţia de a utiliza serviciile oferite de o anumită agenţie turistică, în următoare lună. Ponderea acestora este:

$$\bar{x} = \frac{m}{n} = \frac{75}{300} = 0{,}25 \qquad (25\%)$$

2. Atunci când vrem să combinăm rezultatele obţinute pentru diverse grupe sau componente ale unei colectivităţi (sub forma mediilor aritmetice parţiale),

Scanned with OKEN Scanner

putem calcula un nivel mediu pe întreaga populație, ca medie aritmetică ponderată a mediilor parțiale. Rolul de ponderi îl vor juca volumele grupelor/componentelor:

$$\bar{\bar{x}} = \frac{\sum_{k=1}^{m} \bar{x}_k n_k}{\sum_{k=1}^{m} n_k} \qquad (3.50.)$$

unde:

$\bar{x}_k$ este media grupei k ($k = \overline{1,m}$)

n_k este volumul grupei k($k = \overline{1,m}$)

Exemplul 3.26. Despre patru filiale ale unei companii se cunosc: productivitatea medie a muncii (buc./pers.) și numărul pesonalului productive (Tabelul 3.9.):

Tabelul 3.9.

Filiala	Productivitatea medie a muncii (buc./pers.) ($\bar{x}_k$)	Număr de personal productiv (n_k) (pers.)
A	125	30
B	90	22
C	112	45
D	120	37
Total	–	134

Productivitatea medie a muncii pe întreaga companie este:

$$\bar{\bar{x}} = \frac{\sum_{k=1}^{m} \bar{x}_k n_k}{\sum_{k=1}^{m} n_k} = \frac{125 \cdot 30 + 90 \cdot 22 + 112 \cdot 45 + 120 \cdot 37}{134} = \frac{15210}{134} = 113{,}5 \approx 114 \text{ buc./pers}$$

La fel, media aritmetică la nivelul întregii colectivități se poate determina și pe baza frecvențelor relative.

$$\bar{\bar{x}} = \frac{\sum_{k=1}^{m} \bar{x}_k n_k^*}{\sum_{k=1}^{m} n_k^*} \qquad (3.51.)$$

Scanned with OKEN Scanner

unde $n_k^* = \frac{n_k}{\sum_{k=1}^{m} n_k} = \frac{n_k}{n}$ (3.52.)

Exemplul 3.27. Despre cele două filiale ale unei societăți comerciale cu activitate în domeniul serviciilor publice, se cunosc: salariul mediu în filiala A: 1200 RON/pers., salariul mediu în filiala B: 1820 RON/pers. Știind că în filiala A lucrează 63% din personalul societății comerciale, să se determine salariul mediu pe total.

$\bar{x}_A = 1200$ RON/pers.

$\bar{x}_B = 1820$ RON/pers.

$$n_A^* = \frac{n_A}{n_A + n_B} = \frac{n_A}{n} = 0{,}63$$

$$n_B^* = \frac{n_B}{n_A + n_B} = \frac{n_B}{n} = 1 - n_A^* = 0{,}37$$

$$\bar{x} = \frac{\sum_{k=A,B} \bar{x}_k n_k^*}{\sum_{k=A,B} n_k^*} = \frac{\bar{x}_A \cdot n_A^* + \bar{x}_B \cdot n_B^*}{n_A^* + n_B^*} = 1200 \cdot 0{,}63 + 1820 \cdot 0{,}37 = 1422 \text{ RON/pers.}$$

3. O altă modalitate de a măsura tendința centrală a unei serii este o ***valoare numerică centrală***, situată la distanță egală față de valorile extreme ale seriei. Ea se determină ca medie aritmetică simplă între valoarea minimă și valoarea maximă a seriei:

$$VC = \frac{x_{\min} + x_{\max}}{2} \quad (3.53.)$$

Exemplul 3.28. Pentru seria simplă: 7, 2, 6, 7, 10, 3 valorile extreme sunt: $x_{min} = 2$; $x_{max} = 10$; valoarea numerică centrală este:

$$VC = \frac{x_{\min} + x_{\max}}{2} = \frac{2+10}{2} = 6$$

3.2.5. Alți indicatori de poziție

Indicatorii de poziție, din rândul cărora face parte și mediana, sunt folosiți pentru descrierea poziției unei anumite observații în raport cu întregul set de date. Indicatorii de poziție se numesc generic ***cuantile***.

Scanned with OKEN Scanner

Cuantilele sunt valori care împart o serie ordonată crescător în părți egale. Cuantilele cel mai des utilizate sunt:

- mediana
- cuartilele
- decilele
- percentilele
- ***Cuartilele*** sunt cuantile de ordinul 4. Ele sunt valori ale caracteristicii care împart seria ordonată în patru părți egale. Cuartilele sunt în număr de trei: prima cuartilă (numită și *cuartila inferioară*, notată Q_1), a doua cuartilă (care este și mediana: $Q_2 = Me$) și a treia cuartilă (numită și *cuartila superioară*, notată Q_3).

Cuartila inferioară este valoarea caracteristicii despre care se poate afirma că un sfert din termenii seriei sunt mai mici decât ea și trei sferturi mai mari (Figura 3.4.).

Cuartila superioară este valoarea caracteristicii despre care se poate spune că trei sferturi din termenii seriei sunt mai mici decât ea și un sfert mai mari (Figura 3.4.).

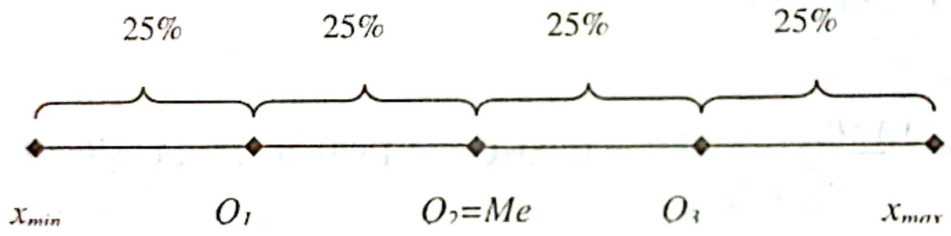


Figura 3.4. Poziționarea cuartilelor

Cuartilele se determină asemănător cu mediana.

- ***Decilele*** sunt valori ale caracteristicii care împart seria ordonată în 10 părți egale. Ele sunt în număr de 9 ($D_1 - D_9$), cea din mijloc (D_5) fiind egală cu mediana ($D_5 = Me$).
- ***Percentilele*** sunt valori ale caracteristicii care împart o serie de valori ordonate în 100 de părți egale. Pentru o serie de date există 99 de percentile ($P_1 - P_{99}$). Între două percentile alăturate se plasează 1% din termenii seriei.

Percentila de rangul (k) este valoarea variabilei pentru care k% din termenii seriei sunt mai mici decât ea și ($100 - k$)% din termeni – mai mari. Percentila cu rangul 25 este cuartila 1 ($P_{25} = Q_1$). Percentila cu rangul 50 este egală cu cuartila 2, cu mediana sau cu decila 5 ($P_{50} = Q_2 = Me = D_5$). Percentila cu rangul 75 este egală cu cuartila 3 ($P_{75} = Q_3$).

Percentila se determină în cazul seriilor cu asimetrie pronunțată. Dacă asimetria este ușoară, atunci este suficient calculul cuartilelor.

Scanned with OKEN Scanner

Pentru determinarea percentilei de rangul k (P_k), în cazul unei serii de repartiție de frecvențe, se parcurg următorii pași:

PAS 1: Se cumulează crescător frecvențele.

PAS 2: Se determină poziția (locul) percentilei în serie.

$$Poz_{Pk} = \frac{k\left(\sum n_i + 1\right)}{100} = \frac{k(n+1)}{100} \qquad (3.54.)$$

unde n este volumul colectivității.

PAS 3: a) dacă seria este pe variante, atunci (P_k) va fi prima variantă a cărei frecvență cumulată este superioară poziției percentilei.

b) dacă seria este pe intervale de variație, atunci se determină intervalul percentilic, ca fiind primul interval a cărui frecvență cumulată este mai mare decât poziția percentilei; se trece apoi la pasul 4.

PAS 4: se calculează valoarea percentilei prin aplicarea unei formule de interpolare pe intervalul percentilic determinat:

$$P_k = x_{\inf(P_k)} + h_{P_k} \cdot \frac{Poz\, P_k - Fc_{(P_k-1)}}{n_{P_k}} \qquad (3.55.)$$

unde: $x_{\inf(P_k)}$ este limita inferioară a intervalului percentilic;

h_{P_k} este mărimea intervalului percentilic;

$Poz\, P_k$ este locul percentilei în serie;

n_{P_k} este frecvența intervalului percentilic;

$Fc_{(P_k-1)}$ frecvența cumulată a intervalului ante-percentilic.

Exemplul 3.29. Aplicând pașii 1-4 vom determina percentilele cu rangul 25 și 75 (sau cuartila inferioară și superioară), pe datele din Tabelul 3.3.

1. Se cumulează crescător frecvențele (vezi Tabelul 3.3., coloana 4).
2. Se determină poziția cuartilei:

$$Poz_{Q_1} = \frac{\sum n_i + 1}{4} = \frac{n+1}{4} = \frac{201}{4} = 50{,}25$$

3. Se identifică intervalul cuartilic: primul interval a cărui frecvență cumulată este mai mare decât Poz_{Q_1}: $Q_1 \in (8-12)$.
4. Se calculează cuartila Q_1 cu relația:

$$Q_1 = x_{\inf(Q_1)} + h_{Q_1} \cdot \frac{Poz\, Q_1 - Fc_{(Q_1-1)}}{n_{Q_1}} =$$

$$8 + 4 \cdot \frac{50{,}25 - 46}{80} = 8{,}21 \text{ ani}$$

Scanned with OKEN Scanner

Asemănător se determină cuartila superioară (Q_3):

$$Poz_{Q_3} = \frac{3\left(\sum n_i + 1\right)}{4} = 150{,}75$$

$$Q_3 \in (12-16)$$

$$Q_3 = x_{\inf(Q_3)} + h_{Q_3} \cdot \frac{Poz\,Q_3 - Fc_{(Q_3-1)}}{n_{Q_3}} =$$

$$12 + 4 \cdot \frac{150{,}75 - 126}{37} = 14{,}68 \text{ ani}$$

Aproximativ 25% dintre agenţii economici au o vechime în activitate sub 8,2 ani şi 75% – mai mare. 75% dintre agenţii economici funcţionează de mai puţin de 14,68 ani şi 25% – peste.

- ***Semicuartila*** este valoarea numerică determinată ca medie aritmetică simplă a cuartilelor Q_1 şi Q_3. Ea este situată la distanţe egale faţă de ambele cuartile extreme.

$$SQ = \frac{Q_1 + Q_3}{2} \qquad (3.56.)$$

3.2.6. Analiza comparativă a principalilor indicatori ai tendinţei centrale

Atunci când analizăm tendinţa centrală a unei serii de date, se pune problema caracterizării acesteia printr-un indicator cât mai potrivit. Pentru alegerea acestuia, dintre cei trei indicatori fundamentali ai tendinţei centrale, trebuie să analizăm natura, tipul seriei. Atunci când seria prezintă un număr limitat de variante, modul este deosebit de util. De exemplu, atunci când vrem să studiem nivelul cererii pentru un produs care se fabrică în câteva mărimi standard (articole de îmbrăcăminte, încălţăminte etc.). Însă atunci când seria prezintă un număr mare de variante, fiind foarte posibil ca fiecare variantă să apară doar o singură dată, modul nu oferă prea multe informaţii folositoare. Într-o asemenea situaţie, se recomandă mai degrabă împărţirea colectivităţii în clase/grupe după variaţia caracteristicii urmărite şi determinarea clasei/grupei modale. Pe de altă parte, dacă seria de date este de natură calitativă, atunci se impune determinarea modului în defavoarea mediei şi medianei. (De exemplu, în ştiinţele comportamentale). Modul are şi avantajul că se determină foarte uşor, mai mult prin observare decât prin calcul. De aceea el este indicat a se utiliza atunci când se doreşte o estimare într-un timp foarte scurt, operativ, sau atunci când interesează doar varianta cu frecvenţa maximă de apariţie. De asemenea, utilizarea modului este indicată atunci când se doreşte obţinerea unei estimări rapide a mediei şi a medianei, în condiţiile în care distribuţia de valori este unimodală şi aproximativ simetrică. În caz contrar şi dacă

Scanned with OKEN Scanner

datele sunt de natură cantitativă, este posibil şi chiar indicat să se determine toţi cei 3 indicatori fundamentali ai tendinţei centrale.

Desigur, alegerea unuia dintre cei 3 indicatori ai tendinţei centrale nu se poate justifica doar prin uşurinţa calculelor acestora, în condiţiile modernizării tehnicii de calcul şi a facilitării accesului la aceasta. Un alt criteriu ar putea fi publicul-ţintă, căruia i se adresează rezultatul studiului. Majoritatea persoanelor înţeleg mai uşor conceptul de „medie aritmetică", în timp ce mediana este mai puţin accesibilă înţelegerii publicului larg. Dacă însă determinăm toţi cei 3 indicatori ai tendinţei centrale, prin compararea lor putem aprecia asimetria sau simetria distribuţiei.

Am văzut că dacă într-o serie există valori extreme aberante, media va fi influenţată într-o mai mare măsură de aceste valori, de aceea, în acest caz, ar fi mai indicat calculul medianei. De asemenea, mediana se dovedeşte utilă şi într-o serie de date ordinale, în timp ce media – nu.

În cazul medianei nu este necesar, pentru determinarea ei, să cunoaştem toate variantele numerice sau nenumerice ale caracteristicii, ci doar să existe o relaţie de ordine între unităţile statistice după valorile variabilei urmărite.

Media este afectată în mai mică măsură de fluctuaţiile de selecţie decât mediana, prezintă o mai mare stabilitate şi se pretează mai uşor calculelor algebrice decât mediana.

Analizând relaţia de ordine dintre cei 3 indicatori ai tendinţei centrale, pentru o distribuţie cu tendinţă de normalitate pot apărea următoarele cazuri:

- în cazul unei distribuţii simetrice unimodale, valorile celor trei indicatori ai tendinţei centrale sunt egale (coincid) [Figura 3.5. a)]:
$$\bar{x} = Me = Mo \qquad (3.57.)$$
- când însă distribuţia are mai multe valori la unul din capetele sale (spre valorile mici sau mari ale variabilei), atunci distribuţia este asimetrică. În figura 3.5. b), aşa cum reiese din forma poligonului frecvenţelor*), distribuţia este pozitiv asimetrică (curba este aplecată spre stânga, coada prelungă fiind spre dreapta), adică în serie există o pondere mai ridicată a valorilor mari ale caracteristicii (mai mari ca modul). În acest caz, media fiind influenţată de valorile extreme, va avea cea mai mare valoare dintre cei 3 indicatori, urmată de mediană şi de mod:
$$\bar{x} > Me > Mo \qquad (3.58.)$$
- dacă poligonul frecvenţelor relative seamănă cu o curbă aplecată spre dreapta, cu coada mai lungă spre stânga (spre valorile mici ale caracteristicii) [Figura 3.5. c)] atunci distribuţia este negativ asimetrică; în acest caz, media este mai mică decât mediana şi modul:
$$\bar{x} < Me < Mo \qquad (3.59.)$$

*) Poligonul frecvenţelor construit pentru un set foarte mare de date tinde să arate ca o curbă, nu ca o linie frântă.

În serie există o pondere mai ridicată a valorilor mici ale caracteristicii (adică mai mici decât modul). Valorile extreme au o influență mai mare asupra mediei decât asupra medianei. Cu cât diferența dintre valorile celor 3 indicatori ai tendinței centrale este mai mare, cu atât distribuția prezintă o asimetrie mai pronunțată.

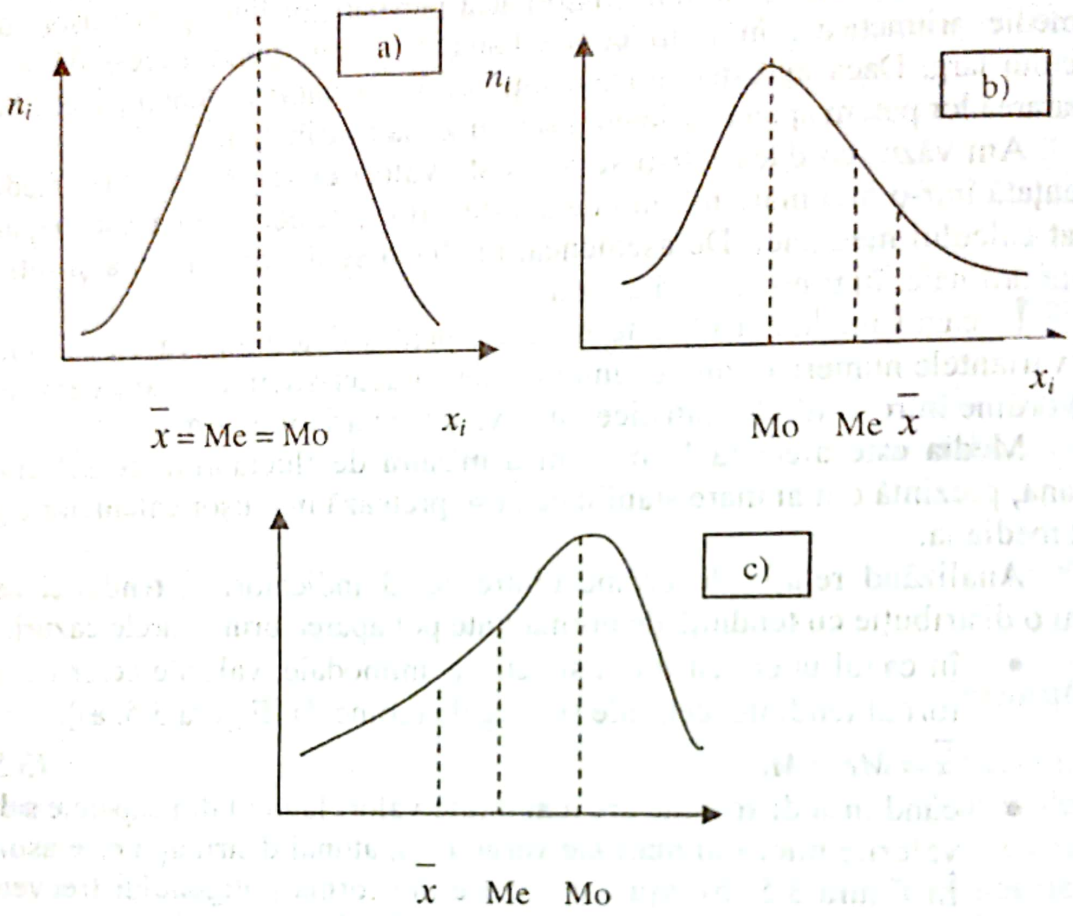


Figura 3.5. Distribuții cu diferite tipuri de asimetrie: a) simetrică; b) pozitiv asimetrică; c) negativ asimetrică

Așa cum am arătat, una dintre cele mai importante funcții ale statisticii descriptive este de a realiza o inferență a rezultatelor de la nivelul eșantionului la nivelul colectivității generale. Media eșantionului poate realiza o bună estimare a mediei colectivității totale.

Rezumarea informațiilor care descriu tendința centrală a unei serii (și nu numai) se realizează în mod grafic cu ajutorul diagramei *Box-and-Whisker Plot*. Ea sintetizează nu numai informațiile referitoare la tendința centrală, ci și cele privind dispersarea valorilor și la forma distribuției. Ea a fost propusă prima dată de Tukey în 1977, dar de atunci au mai fost și alte propuneri de modificare a formei sale.

Scanned with OKEN Scanner

Cu ajutorul diagramei se pot examina datele în vederea identificării unor posibile erori. Într-o diagramă *Box-and-Whisker Plot* se redau următoarele informații:

- valoarea minimă (x_{min});
- valoarea maximă (x_{max});
- cuartila inferioară (Q_1);
- cuartila superioară (Q_3);
- mediana (Me);
- media aritmetică ($\bar{x}$).

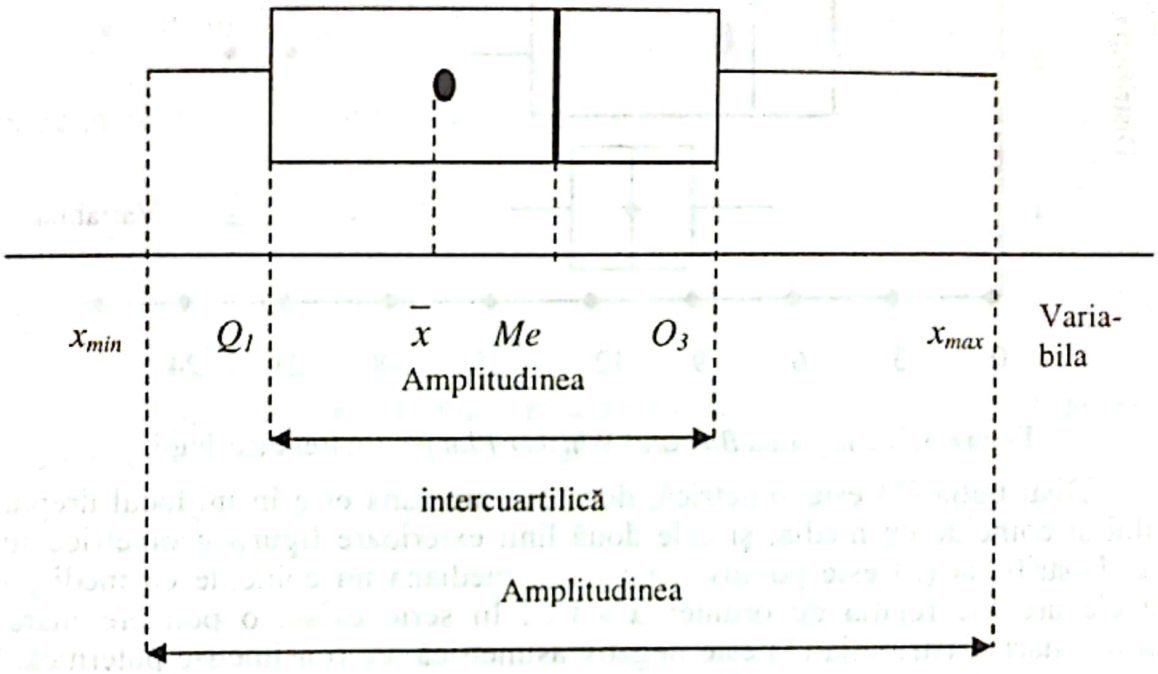


Figura 3.6. Diagrama *Box-and-Whisker Plot*

Diagrama este alcătuită dintr-un dreptunghi şi câteva linii. Ea se poate trasa pe orizontală sau pe verticală. Laturile mici ale dreptunghiului marchează pe axă cuartilele inferioară şi superioară, linia din interiorul dreptunghiului va indica pe axă mediana, punctul din interiorul dreptunghiului va marca media aritmetică. Cele două linii exterioare dreptunghiului se duc până în dreptul valorilor minime, respectiv maxime ale seriei. Lungimea dreptunghiului reprezintă amplitudinea intercuartilică ($Q_3 - Q_1$), iar lungimea totală a graficului (de la un capăt al primei linii exterioare dreptunghiului până la capătul final al celei de-a doua linii) desemnează amplitudinea totală a variației. Ambii indicatori redau informații referitoare la dispersarea valorilor față de medie sau unele față de altele. Poziția medianei față de medie indică dacă datele sunt distribuite simetric sau nu.

Dacă valorile extreme sunt şi aberante, atunci acestea se reprezintă pe grafic în afara dreptunghiului prin unul sau două asteriscuri.

Scanned with OKEN Scanner

Exemplul 3.30. Se consideră trei distribuţii pentru care s-au întocmit diagramele *Box-and-Whisker Plot* (Figura 3.7.):

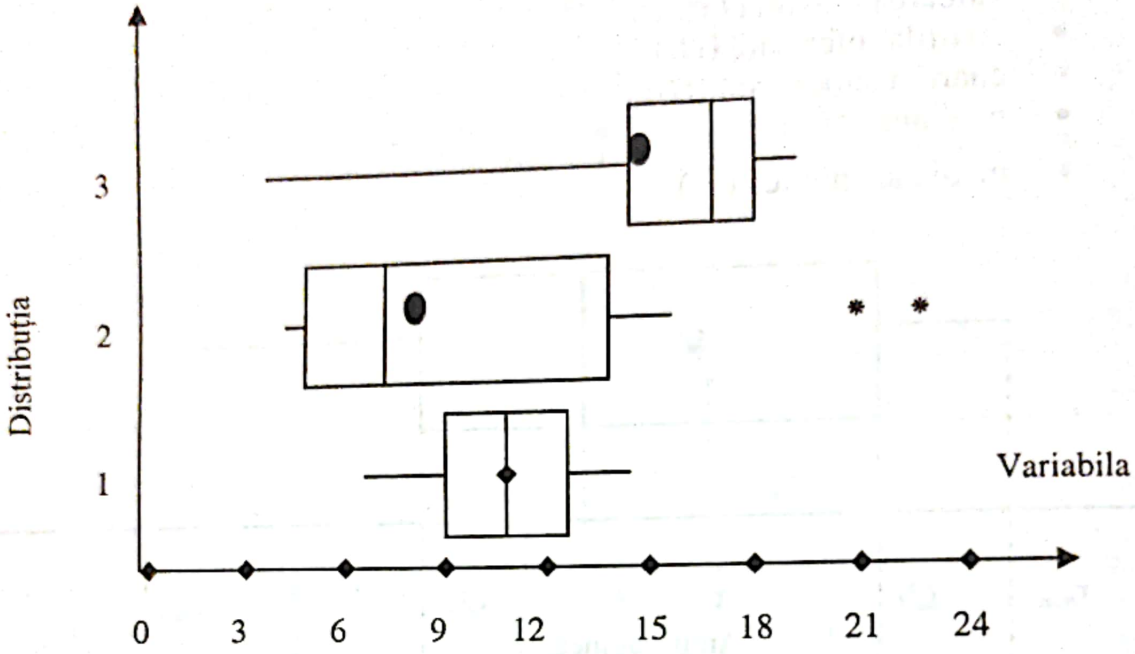


Figura 3.7. Diagrama *Box-and-Whisker Plot* pentru trei distribuţii

Distribuţia (1) este simetrică, deoarece mediana este în mijlocul dreptunghiului şi coincide cu media, şi cele două linii exterioare figurii geometrice sunt egale. Distribuţia (2) este pozitiv asimetrică, mediana nu coincide cu media, iar între ele are loc relaţia de ordine: $x > Me$. În serie există o pondere mare a valorilor mari. Distribuţia (3) este negativ asimetrică, cu o asimetrie puternică. În acest caz, $x < Me$, ceea ce indică faptul că există o pondere ridicată a valorilor mici ale caracteristicii.

3.2.7. Medii cu aplicaţii speciale utilizate în statistică

Există unele tipuri mai deosebite de medii, utilizate ocazional în descrierea seriilor de date, în anumite situaţii mai speciale.

1. ***Media armonică*** (x_h) este valoarea care, înlocuind fiecare termen al seriei, nu modifică suma inverselor acestora. Se mai numeşte „inversa mediei aritmetice a inverselor termenilor seriei". Se determină, ca şi media aritmetică, în variantă simplă (pentru o serie simplă) şi în variantă ponderată (pentru distribuţii de frecvenţe absolute sau relative).

- pentru o serie simplă, de forma $(x_i), i = \overline{1,n}$, suma inverselor termenilor se scrie:

$$\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} = \sum_{i=1}^{n} \frac{1}{x_i} \qquad (3.60.)$$

Înlocuind fiecare termen al seriei cu media armonică a acesteia, obținem:

$$\frac{1}{\bar{x}_h}+\frac{1}{\bar{x}_h}+\ldots+\frac{1}{\bar{x}_h}=\sum_{i=1}^{n}\frac{1}{x_i} \qquad (3.61.)$$

$$\frac{n}{\bar{x}_h}=\sum_{i=1}^{n}\frac{1}{x_i} \quad \Rightarrow \bar{x}_h=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \text{ (medie armonică simplă)} \qquad (3.62.)$$

- Pentru o serie de distribuție de frecvențe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i=\overline{1,r}$ se

obține în mod asemănător relația:

$$\bar{x}_h=\frac{\sum_{i=1}^{r} n_i}{\sum_{i=1}^{r}\frac{1}{x_i}n_i} \qquad (3.63.)$$

Relația (3.63.) reprezintă formula de calcul a mediei armonice ponderate, utilizate pentru distribuții de frecvențe absolute.

$$\bar{x}_h=\frac{\sum_{i=1}^{r} n_i^*}{\sum_{i=1}^{r}\frac{1}{x_i}n_i^*} \qquad (3.64.)$$

Relația (3.64.) reprezintă formula de calcul a mediei armonice ponderate utilizate pentru distribuții de frecvențe relative, exprimate în coeficienți sau în procente.

Media armonică se folosește în cazul seriilor statistice pronunțat asimetrice în care predomină valorile mici, distribuția fiind în formă de „J" întors [Figura 3.8. a)].

Scanned with OKEN Scanner

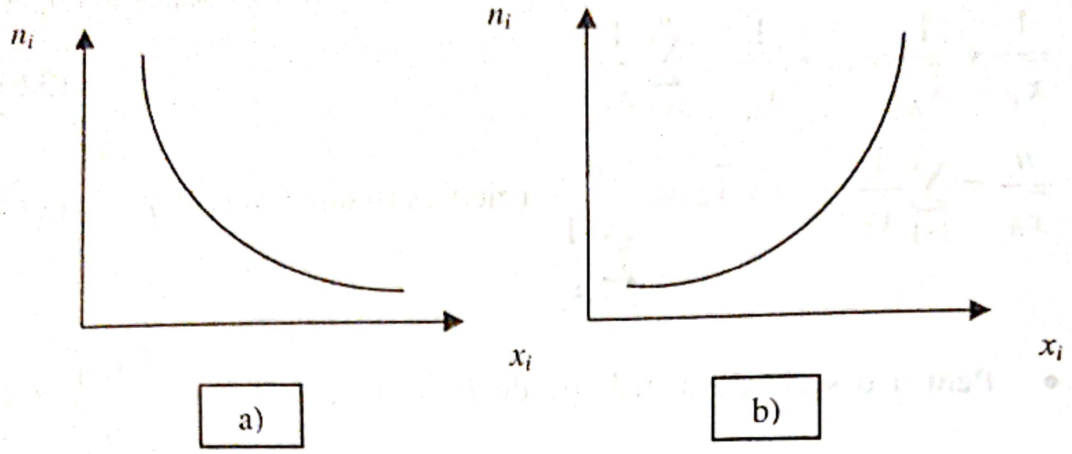


Figura 3.8. Serii statistice pronunțat asimetrice cu predominanță a valorilor: a) mici; b) mari

În statistică, media armonică se regăseşte în calculul valorilor medii ale unor caracteristici exprimate ca raport. Media armonică este întotdeauna mai mică decât media aritmetică, dacă ele sunt calculate pentru o aceeaşi serie de date.

Exemplul 3.31. Un agent economic a avut într-o lună o productivitate a muncii de $w_1 = 5$ unități monetare/pers., în timp ce un alt agent economic a lucrat în aceeaşi lună cu o productivitate de $w_2 = 9$ unități monetare/pers. Ştiind că cei doi agenți economici au realizat producții de $Q_1 = 150$, respectiv $Q_2 = 180$ unități monetare, să se determine productivitatea medie a muncii pe ansamblul celor doi agenți economici.

$$\overline{W} = \frac{\sum Q_i}{\sum N_i} = \frac{\sum Q_i}{\sum \frac{Q_i}{w_i}} = \frac{Q_1 + Q_2}{\frac{Q_1}{w_1} + \frac{Q_2}{w_2}} = \frac{150+180}{\frac{150}{5} + \frac{180}{9}} = \frac{330}{30+20} = \frac{330}{50} = 6{,}6 \text{ u.m./pers.}$$

unde N_i este numărul de personal pentru agentul economic i.

Productivitatea medie a muncii s-a calculat ca medie armonică ponderată a productivităților fiecărui agent economic. Rolul de ponderi l-au jucat producțiile celor doi agenți.

2. Media pătratică ($\bar{x}_p$) este valoarea care, înlocuind fiecare termen al seriei, nu modifică suma pătratelor acestora. Ea reprezintă rădăcina pătrată a raportului între suma pătratelor termenilor şi numărul acestora. Se determină în variantă simplă (pentru serii simple) şi în variantă ponderată (pentru distribuții de frecvențe).

- pentru o serie simplă, de forma $(x_i), i = \overline{1,n}$, suma pătratelor termenilor se scrie:

$$x_1^2 + x_2^2 + ... + x_n^2 = \sum_{i=1}^{n} x_i^2 \qquad (3.65.)$$

Înlocuind fiecare termen al seriei cu media pătratică a acesteia, obținem:

$$\bar{x}_p^2 + \bar{x}_p^2 + ... + \bar{x}_p^2 = \sum_{i=1}^{n} x_i^2 \qquad (3.66.)$$

$$n \cdot \bar{x}_p^2 = \sum_{i=1}^{n} x_i^2 \Rightarrow \bar{x}_p = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \text{ (medie pătratică simplă)} \qquad (3.67.)$$

- Pentru o serie de distribuție de frecvențe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i = \overline{1,r}$ se

obține în mod asemănător relația:

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{r} x_i^2 n_i}{\sum_{i=1}^{r} n_i}} \qquad (3.68.)$$

Relația (3.68.) reprezintă formula de calcul a mediei pătratice ponderate, utilizată pentru distribuții de frecvențe absolute.

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{r} x_i^2 n_i^*}{\sum_{i=1}^{r} n_i^*}} \qquad (3.69.)$$

Relația (3.69.) reprezintă formula de calcul a mediei pătratice ponderate, utilizată pentru distribuții de frecvențe relative.

Media pătratică se utilizează atunci când într-o serie predomină valorile mari, seria find pronunțat asimetrică spre acestea [în formă de „J" – Figura 3.8. b)]. Prin ridicarea la pătrat, cu ajutorul mediei pătratice se pun în valoare nivelurile mai ridicate ale caracteristicii urmărite. În statistică, acest tip de medie se folosește la determinarea unor indicatori ai variației (de exemplu, abaterea medie pătratică). Media pătratică este întotdeauna superioară ca valoare atât mediei aritmetice, cât și mediei armonice (pentru o aceeași serie de date).

3. ***Media geometrică*** ($\bar{x}_g$) este valoarea care, înlocuind fiecare termen al seriei, nu modifică produsul acestora. Ea se definește ca rădăcina de ordinul n a

Scanned with OKEN Scanner

produsului a n termeni ai seriei. Se determină în variantă simplă (pentru serii simple) şi în variantă ponderată (pentru distribuţii de frecvenţe).

- pentru o serie simplă, de forma $(x_i), i=\overline{1,n}$, produsul termenilor seriei se scrie:

$$x_1 \cdot x_2 \cdot \ldots \cdot x_n = \prod_{i=1}^{n} x_i \qquad (3.70.)$$

Înlocuind fiecare termen al seriei cu media geometrică a acesteia, obţinem:

$$\bar{x}_g \cdot \bar{x}_g \cdot \ldots \cdot \bar{x}_g = \prod_{i=1}^{n} x_i \qquad (3.71.)$$

$$\bar{x}_g^{n} = \prod_{i=1}^{n} x_i \Rightarrow \bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i} \text{ (medie geometrică simplă)} \qquad (3.72.)$$

- Pentru o serie de distribuţie de frecvenţe, de forma: $\begin{pmatrix} x_i \\ n_i \end{pmatrix}, i=\overline{1,r}$ se obţine în mod asemănător relaţia:

$$\bar{x}_g = \sum \sqrt[n_i]{\prod_{i=1}^{r} x_i^{n_i}} \text{ (medie geometrică ponderată)} \qquad (3.73.)$$

Relaţia (3.73.) se utilizează în cazul distribuţiilor de frecvenţe absolute. Dacă înlocuim frecvenţele absolute n_i cu cele relative (n_i^*), se obţine o formulă de calcul asemănătoare. Pentru determinarea mediei geometrice, se foloseşte logaritmul:

$$\log\left(\bar{x}_g\right) = \log\left(\prod_{i=1}^{n} x_i\right)^{1/n} = \frac{\log\left(\prod_{i=1}^{n} x_i\right)}{n} = \frac{\sum_{i=1}^{n} \log x_i}{n} \qquad (3.74.)$$

$$\bar{x}_g = anti\log\left(\frac{\sum_{i=1}^{n} \log x_i}{n}\right) \qquad (3.75.)$$

Din relaţia (3.74.) se observă că logaritmul mediei geometrice este egal cu media aritmetică a logaritmilor termenilor seriei.

Media geometrică are dezavantajul că nu se poate folosi atunci când variabila are valori negative sau nule (în ultimul caz, media este şi ea nulă). Acest tip de medie se utilizează atunci când termenii seriei au forma unor rapoarte

Scanned with OKEN Scanner

(indici), ce reflectă modificările în timp ale unui fenomen (eventual modificări în ritm aproximativ constant). În statistică, media geometrică se foloseşte în determinarea unor indici medii de dinamică.

Exemplul 3.32. Indicii de dinamică ai producţiei realizate de o societate comercială în perioada 2000-2004 sunt (Tabelul 3.10.):

Tabelul 3.10.

Ani	Indicele de dinamică a producţiei faţă de anul precedent (%) ($x_{t/t-1}$)
2001	125
2002	119
2003	137
2004	142

Pentru a determina indicele mediu de dinamică al producţiei, vom utiliza media cronologică, întrucât termenii seriei sunt rezultatul unor rapoarte (între producţia anului curent şi producţia anului precedent).

$$\bar{I}_{2000-2004} = \sqrt[4]{\prod_{t=2001}^{2004} x_{t/t-1}} = \sqrt[4]{x_{2001/2000} \cdot x_{2002/2001} \cdot x_{2003/2002} \cdot x_{2004/2003}}$$

$$\bar{I}_{2000-2004} = \sqrt[4]{125 \cdot 119 \cdot 137 \cdot 142} = 130{,}43\%$$

Între media aritmetică, armonică, pătratică şi geometrică, calculate pentru aceeaşi serie de date, există următoarea relaţie de ordine:

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_p \qquad (3.76.)$$

Oricare din aceste tipuri de medii se găsesc între valoarea minimă şi valoarea maximă a caracteristicii.

Un alt tip de medie utilizat în statistică este media cronologică, ce va fi detaliată în capitolul 6.

3.3. Analiza statistică a variabilităţii

Încă din capitolul 1 am văzut că statistica studiază fenomene de masă, adică fenomene care se produc la numeroase unităţi statistice şi variază ca nivel de la o unitate la alta. Aşadar, variabilitatea este o trăsătură esenţială a fenomenelor şi proceselor care formează obiectul de studiu al statisticii, este, totodată, şi un concept statistic fundamental, un aspect de bază urmărit în orice proces de investigaţie statistică şi care prezintă un interes deosebit pentu orice cercetător.

Dacă, de exemplu, rata şomajului în România, în anul 2005 ar avea aproximativ aceleaşi valori în toate localităţile ţării, atunci acest fenomen nu ar prezenta interes pentru un studiu statistic aprofundat, limitându-ne la a determina un nivel mediu al fenomenului. Cum însă nu se întâmplă astfel, existând localităţi cu rate

Scanned with OKEN Scanner

foarte ridicate ale şomajului (precum cele din nord-estul ţării, din regiuni miniere în care au fost închise unele mine de exploatare etc.) şi, dimpotrivă, localităţi cu rate mai scăzute ale şomajului (precum Bucureşti), fenomenul trebuie studiat pentru a explica gradul de dispersare a valorilor acestuia, factorii sub influenţa cărora iau naştere aceste diferenţe, în vederea adoptării unor decizii de remediere a situaţiilor de evoluţie nefavorabilă a fenomenului. Aşadar, indicatorii de măsurare a variabilităţii completează informaţiile aduse de indicatorii descriptivi tendinţei centrale, pentru o analiză mai aprofundată a unui set de date.

Două distribuţii pot manifesta tendinţe centrale asemănătoare, dar pot avea, în acelaşi timp, variaţii diferite, grade diferite de împrăştiere a valorilor [Figura 3.9. a)]. Distribuţiile normale A şi B din Figura 3.9. a) au o aceeaşi tendinţă centrală exprimată prin media aritmetică, dar distribuţia A prezintă o variaţie mai mică a valorilor decât distribuţia B.

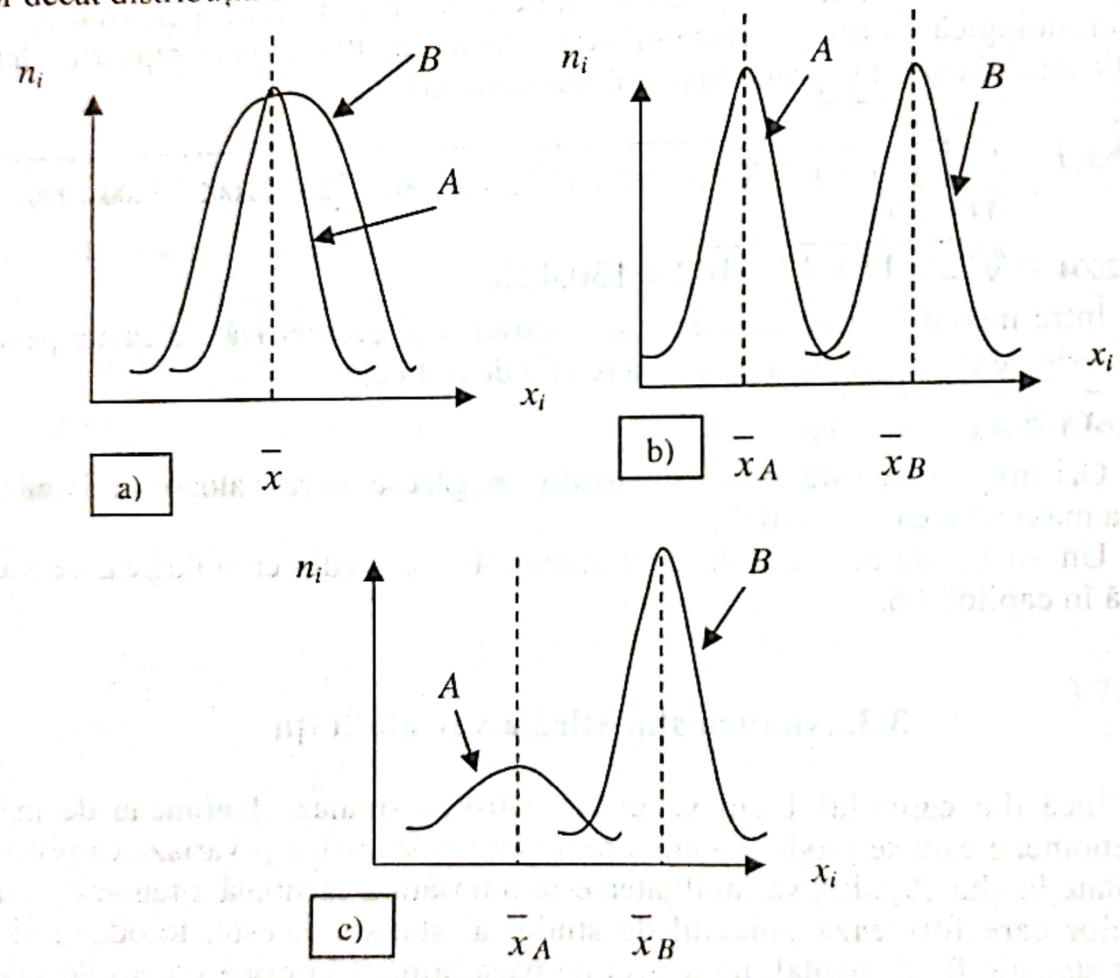


Figura 3.9. Serii statistice cu tendinţă centrală şi variaţie asemănătoare sau diferite: a) serii cu aceeaşi tendinţă centrală şi variaţie diferită; b) serii cu acelaşi grad de variaţie şi tendinţă centrală diferită; c) serii cu tendinţă centrală şi variaţie diferite.

Scanned with OKEN Scanner

În alte situații, două serii pot avea tendințe centrale diferite, dar grade identice de dispersare a valorilor [Figura 3.9. b)]. În sfârșit, două serii pot diferi atât prin tendința centrală, cât și prin variabilitate [Figura 3.9. c)].

Este important, pentru analiza unui set de date, să aflăm la ce nivel mediu se situează caracteristica la care se referă datele, dar aceste informații trebuie completate cu unele privind gradul de împrăștiere a acestor date. Dacă ele sunt prea dispersate unele față de altele sau față de medie, atunci nivelul mediu, ca expresie a tendinței centrale, nu ar mai fi reprezentativ pentru seria din care s-a calculat.

Variabilitatea este caracterizată și măsurată printr-un sistem de indicatori, care se pot clasifica după următoarele criterii:

– după nivelul la care se calculează și aria de cuprindere, avem:
- indicatori simpli ai variației (calculați la nivelul unei unități statistice);
- indicatori sintetici ai variației (calculați la nivelul întregii colectivități sau al unei grupe a acesteia).

– după forma de exprimare, pot fi:
- indicatori absoluți (exprimă variația în unități concrete de măsură);
- indicatori relativi (exprimă variația în procente sau coeficienți).

Conceptul de variabilitate este utilizat în fundamentarea multor decizii zilnice, chiar dacă variabilitatea este exprimată numeric sau nu.

Exemplul 3.33. Să presupunem că avem două caracteristici, X și Y, cu următoarele valori:

$X = \{7,8,9\}$

$Y = \{2,8,14\}$

Se observă că ambele caracteristici au media aritmetică egală cu opt, dar că variabila Y are valorile mai dispersate unele față de altele, decât X.

3.3.1. Indicatorii simpli de măsurare a variabilității

Indicatorii simpli ai variației se pot exprima în formă absolută și relativă.

I. Indicatorii simpli absoluți ai variației compară sub formă de diferență mărimea a doi termeni ai seriei sau mărimea unui termen al seriei cu nivelul mediu al acesteia. Se exprimă în unități concrete de măsură.

- ***Amplitudinea absolută a variației*** este diferența dintre valoarea maximă și valoarea minimă a unei serii:

$$A = x_{max} - x_{min} \qquad (3.77.)$$

Amplitudinea absolută se exprimă în aceeași unitate de măsură ca și caracteristica pentru care s-a calculat.

Exemplul 3.34. Pentru un eșantion de 5 valori: 7, 9, 12, 10, 7, amplitudinea este $A = 12 - 7 = 5$ u.m. Rezultatul ne indică faptul că cele 5 valori ale caracteristicii urmărite se află la o distanță de cel mult 5 u.m. unele față de altele.

Scanned with OKEN Scanner

Probleme mai deosebite apar atunci când cele două valori extreme sunt şi valori aberante, aflate la distanţă considerabilă faţă de ceilalţi termeni ai seriei. În acest caz, amplitudinea nu mai prezintă o valoare semnificativă, fiind mai degrabă indicat a se reţine nivelul inferior şi superior ale caracteristicii, decât diferenţa dintre ele. Nivelul indicatorului poate fi comparat numai atunci când se referă la aceeaşi caracteristică, măsurată în aceeaşi unitate de măsură.

Dacă avem o distribuţie de frecvenţe pe intervale de variaţie, atunci amplitudinea se determină scăzând din limita superioară a ultimului interval limita inferioară a primului interval.

Amplitudinea prezintă şi unele dezavantaje:

- este destul de instabilă, pe măsura adăugării de noi date;
- prezintă sensibilitate faţă de valorile extreme;
- nu poate fi folosită în comparaţii, decât pentru serii care se referă la aceeaşi caracteristică (care se exprimă în aceeaşi unitate de măsură);
- dacă într-o distribuţie de frecvenţe pe intervale de valori nu se cunosc limitele inferioară şi superioară, atunci amplitudinea nu poate fi calculată, decât făcând anumite presupuneri pe baza mărimii intervalelor de grupare.

- ***Abaterea (devierea) individuală absolută*** se determină ca diferenţă între valoarea unui termen al seriei şi nivelul mediu al acesteia; şi ea se exprimă, ca şi amplitudinea absolută, în aceeaşi unitate de măsură ca şi caracteristica pentru care se calculează şi nu poate fi folosită în comparaţii între serii care se referă la variabile diferite. Se poate calcula cu relaţia:

$$d_i = x_i - \bar{x}, \text{ cu } i = \overline{1,n} \qquad (3.78.)$$

unde i este unitatea statistică.

Exemplul 3.35. Pentru datele din exemplul (3.34.), să determinăm abaterile individuale absolute ale termenilor seriei faţă de medie (Tabelul 3.11.):

$$\bar{x} = \frac{\sum x_i}{n} = \frac{45}{5} = 9$$

Tabelul 3.11.

x_i	7	7	9	10	12
$d_i = x_i - \bar{x}$	–2	–2	0	1	3

- ***Amplitudinea intercuartilică*** se defineşte ca diferenţă între cuartila superioară şi cea inferioară:

$$A_Q = Q_3 - Q_1 \qquad (3.79.)$$

Arată între ce valori se încadrează jumătate din termenii seriei.

Scanned with OKEN Scanner

- ***Abaterea semiintercuartilică*** reprezintă jumătate din amplitudinea cuartilică:

$$A_{SQ} = \frac{Q_3 - Q_1}{2} \qquad (3.80.)$$

- ***Abaterea maximă pozitivă*** (în expresie absolută) este diferența între valoarea maximă a caracteristicii și valoarea medie a acesteia:

$$d^+_{\max} = x_{\max} - \bar{x} \qquad (3.81.)$$

Exemplul 3.36. În exemplul precedent (3.35.), abaterea maximă pozitivă absolută este:

$$d^+_{\max} = x_{\max} - \bar{x} = 12 - 9 = 3 \text{ u.m.}$$

- ***Abaterea maximă negativă*** (în expresie absolută) este diferența dintre valoarea minimă a caracteristicii și valoarea medie a acesteia:

$$d^-_{\max} = x_{\min} - \bar{x} \qquad (3.82.)$$

Exemplul 3.37. În exemplul 3.35., abaterea maximă negativă absolută este:

$$d^-_{\max} = x_{\min} - x = 7 - 9 = -2 \text{ u.m.}$$

II. Indicatorii simpli relativi ai variației se determină sub formă de raport, de regulă, între un indicator absolut individual al variabilității și media seriei. Se exprimă în procente sau coeficienți, nemaidepinzând de unitatea de măsură a caracteristicii pentru care se determină.

- ***Amplitudinea relativă a variației*** este raportul dintre amplitudinea absolută a variației caracteristicii și nivelul mediu al acesteia.

$$A_{\%} = \frac{A}{\bar{x}} \cdot 100 = \frac{x_{\max} - x_{\min}}{\bar{x}} \cdot 100 \qquad (3.83.)$$

- ***Abaterea individuală relativă*** este raportul dintre abaterea individuală absolută și nivelul mediu al caracteristicii.

$$d_{\%} = \frac{d_i}{\bar{x}} \cdot 100 = \frac{x_i - \bar{x}}{\bar{x}} \cdot 100 \qquad (3.84.)$$

- ***Abaterea maximă pozitivă relativă*** este raportul dintre abaterea maximă pozitivă absolută și nivelul mediu al caracteristicii:

$$d^+_{\max\%} = \frac{d^+_{\max}}{\bar{x}} \cdot 100 = \frac{x_{\max} - \bar{x}}{\bar{x}} \cdot 100 \qquad (3.85.)$$

- ***Abaterea maximă negativă relativă*** se determină raportând abaterea maximă negativă absolută la nivelul mediu al seriei:

$$d^-_{\max\%} = \frac{d^-_{\max}}{\bar{x}} \cdot 100 = \frac{x_{\min} - \bar{x}}{\bar{x}} \cdot 100 \qquad (3.86.)$$

3.3.2. Indicatorii sintetici de măsurare a variabilității

Măsoară în mod sintetic abaterea unui termen al seriei față de nivelul mediu al acesteia.

I. Abaterea medie absolută (sau abaterea medie liniară)

Dacă abaterile individuale absolute d_i au valori mai mari, atunci înseamnă că termenii pentru care s-au calculat sunt situați la distanțe mai mari față de medie, în timp ce, dacă aceste abateri au valori mai mici, temenii respectivi sunt mai apropiați ca nivel de media seriei. Abaterile se pot produce, deci, într-un sens sau în altul față de medie, dar am văzut că pe ansamblul întregii colectivități ele se compensează reciproc, astfel că avem:

- pentru o serie simplă:

$$\sum_{i=1}^{n} d_i = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} \left(x_i - \bar{x}\right) = 0 \tag{3.87.}$$

- pentru o serie de frecvențe:

$$\sum_{i=1}^{r} \left(x_i - \bar{x}\right) n_i = 0 \tag{3.88.}$$

Aceste relații sunt în conformitate cu proprietatea (4) a mediei aritmetice.

De aceea, pentru a evita anularea reciprocă a abaterilor individuale absolute a termenilor seriei față de medie, se utilizează valoarea absolută (în modul) a acestor abateri:

$$\sum_{i=1}^{n} \left|d_i\right| = 0 \tag{3.89.}$$

Calculând media aritmetică simplă sau ponderată a abaterilor individuale luate în valoarea absolută, ale termenilor seriei față de media lor, se obține indicatorul abaterea medie absolută (sau liniară), exprimat în unități concrete de măsură, aceleaşi ca şi caracteristica pentru care s-a calculat indicatorul. Acesta arată cu câte unități de măsură concrete s-a modificat în medie un termen al seriei față de nivelul mediu al acesteia. Se determină potrivit relațiilor:

- pentru o serie simplă:

$$\bar{d} = \frac{\sum_{i=1}^{n} \left|x_i - \bar{x}\right|}{n} \tag{3.90.}$$

Scanned with OKEN Scanner

- pentru o serie de frecvenţe absolute:

$$\bar{d} = \frac{\sum_{i=1}^{r} |x_i - \bar{x}| n_i}{\sum_{i=1}^{r} n_i} \qquad (3.91.)$$

- pentru o serie de frecvenţe relative:

$$\bar{d} = \frac{\sum_{i=1}^{r} |x_i - \bar{x}| n_i^*}{\sum_{i=1}^{r} n_i^*} \qquad (3.92.)$$

Exemplul 3.38. Pentru seria simplă din exemplul (3.34.), abaterea medie liniară este:

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} = \frac{2+2+0+1+3}{5} = \frac{8}{5} = 1{,}6 \text{ u.m.}$$

Exemplul 3.39. Pentru distribuţia de frecvenţe din Tabelul 3.3., avem:

$$\bar{d} = \frac{\sum_{i=1}^{r} |x_i - \bar{x}| n_i}{\sum_{i=1}^{r} n_i} = \frac{796{,}32}{200} = 3{,}98 \text{ ani}$$

Calculele intermediare se găsesc în Tabelul 3.12. (coloanele 3, 4):

Tabelul 3.12.

Nr. de ani de existenţă	Nr. de agenţi economici	x_i	d_i	$\lvert x_i - \bar{x}\rvert n_i$	$(x_i - \bar{x})^2 n_i$	x_i^2	$x_i^2 n_i$
0	1	2	3	4	5	6	7
0-4	17	2	–9,16	155,72	1426,395	4	68
4-8	29	6	–5,16	149,64	772,142	36	1044
8-12	80	10	–1,16	92,8	107,648	100	8000
12-16	37	14	2,84	105,08	298,427	196	7252
16-20	27	18	6,84	184,68	1263,211	324	8748
20-24	10	22	10,84	108,4	1175,056	484	4840
Total	200	–	–	796,32	5042,879	–	29.952

Scanned with OKEN Scanner

Abaterea medie liniară prezintă o serie de dezavantaje:

- se pretează mai greu la calculele algebrice;
- depinde de unitatea de măsură a caracteristicii pentru care se calculează, de aceea nu poate fi folosită în compararea variației unor serii diferite, corespunzătoare unor variabile care se exprimă în unități de măsură diferite;
- acordă importanță în egală măsură abaterilor individuale mai mari sau mai mici în valoare absolută, a termenilor seriei față de media lor, cu toate că variația este influențată în primul rând de abaterile mai mari.

II. Dispersia (sau varianța) este unul din cei mai des utilizați indicatori sintetici ai variabilității, calculat ca medie aritmetică simplă sau ponderată a pătratelor abaterilor individuale absolute, ale termenilor seriei față de media lor. Relația de calcul este:

- pentru o serie simplă:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \qquad (3.93.)$$

- pentru o serie de frecvențe absolute:

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i} \qquad (3.94.)$$

- pentru o serie de frecvențe relative:

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i^*}{\sum_{i=1}^{r} n_i^*} \qquad (3.95.)$$

Deşi dispersia nu are unitate de măsură, ea prezintă, totuşi, avantajul că, prin ridicarea la pătrat, acordă o importanță mai mare abaterilor individuale mai mari în valoare absolută față de media seriei:

Exemplul 3.39. Pentru seria simplă din exemplul (3.34.), dispersia va fi:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{18}{5} = 3{,}6$$

Calculele intermediare se găsesc în Tabelul 3.13.

Scanned with OKEN Scanner

Tabelul 3.13.

x_i	7	7	9	10	12	Total 45
$d_i = x_i - \bar{x}$	−2	−2	0	1	3	0
$(x_i - \bar{x})^2$	4	4	0	1	9	18

Exemplul 3.40. Pentru distribuția de frecvențe din Tabelul 3.3., vom determina dispersia:

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i} = \frac{5042{,}879}{200} = 25{,}21$$

Calculele intermediare se regăsesc în Tabelul 3.12., coloanele 3-5.

Prelucrând relația (3.93.), obținem, pentru seria simplă:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i^2 - 2\cdot\bar{x}\cdot x_i + \bar{x}^2\right)}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - 2\cdot\bar{x}\cdot\frac{\sum_{i=1}^{n} x_i}{n} + \frac{n\cdot\bar{x}^2}{n} =$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 = \bar{x}_p^2 - \bar{x}^2 \qquad (3.96.)$$

Asemănător, pentru distribuția de frecvențe, obținem:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n}\left(x_i^2 - 2\cdot\bar{x}\cdot x_i + \bar{x}^2\right) n_i}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} x_i^2 n_i}{\sum_{i=1}^{n} n_i} - 2\cdot\bar{x}\cdot\frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i} + \frac{\bar{x}^2\cdot\sum_{i=1}^{n} n_i}{\sum_{i=1}^{n} n_i} =$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{\sum_{i=1}^{n} n_i} - 2\cdot\bar{x}^2 + \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{\sum_{i=1}^{n} n_i} - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{\sum_{i=1}^{n} n_i} - \left(\frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i}\right)^2 = \bar{x}_p^2 - \bar{x}^2 \qquad (3.97.)$$

Scanned with OKEN Scanner

Notăm cu: $MI_k = \frac{\sum x_i^k}{n}$ momentul inițial de ordinul k pentru o serie simplă, respectiv $MI_k = \frac{\sum x_i^k n_i}{\sum n_i}$ pentru o serie de frecvențe. Rezultă, așadar, din relațiile (3.96.) și (3.97.) că dispersia se poate calcula și ca diferență între momentul inițial de ordinul 2 și pătratul momentului inițial de ordinul 1, sau ca diferență între pătratul mediei pătratice și pătratul mediei aritmetice. Momentul inițial de ordinul 1 este media aritmetică.

Pe lângă momentele inițiale, există și momente centrate, în care se iau în considerare abaterile termenilor seriei față de medie. Astfel, $MC_k = \frac{\sum \left(x_i - \bar{x}\right)^k}{n}$ reprezintă momentul centrat de ordinul k pentru o serie simplă, iar $MC_k = \frac{\sum \left(x_i - \bar{x}\right)^k n_i}{\sum n_i}$ este momentul centrat de ordinul k pentru o distribuție de frecvențe. Înlocuind frecvențele absolute cu cele relative, obținem relațiile de calcul ale momentelor pentru o distribuție de frecvențe relative. Momentul centrat de ordinul 1 este egal cu 0, iar momentul centrat de ordinul 2 este egal cu dispersia.

Exemplul 3.41. Pentru seria simplă din exemplul (3.34.), vom calcula dispersia ca diferență de momente.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2 = \frac{423}{5} - 81 = 3{,}6$$

Am obținut aceeași valoare ca cea calculată prin relația (3.93.).

Exemplul 3.42. Pentru distribuția de frecvențe din Tabelul 3.3., vom determina dispersia ca diferență de momente:

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{\sum_{i=1}^{n} n_i} - \left(\frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i} \right)^2 = \frac{29.952}{200} - \left(\frac{2232}{200} \right)^2 = 25{,}21$$

S-a obținut aceeași valoare ca și prin formula (3.93.).

Scanned with OKEN Scanner

Proprietăţile dispersiei

1. Pentru un şir de valori constante, dispersia este întotdeauna nulă.

Fie o serie simplă de forma $(x_i)_{i=\overline{1,n}}$, cu $x_1 = x_2 = ... = x_n = c = \bar{x}$ (conform proprietăţii (3) a mediei aritmetice).

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (c-c)^2}{n} = 0 \qquad (3.98.)$$

Asemănător se demonstrează această proprietate şi pentru o distribuţie de frecvenţe.

2. Dacă într-o serie (simplă sau de frecvenţe), pentru care s-a calculat media $\bar{x}$ şi dispersia s_x^2, fiecare termen se măreşte sau se micşorează cu o aceeaşi valoare c, atunci dispersia seriei nu se modifică.

- Fie o serie simplă, de forma $(x_i)_{i=\overline{1,n}}$, pentru care s-a calculat media $\bar{x} = \frac{\sum x_i}{n}$ şi dispersia $s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$. Noua serie, provenită din mărirea fiecărui termen al vechii serii cu o constantă c, va fi de forma: $(y_i)_{i=\overline{1,n}}$, unde $y_i = x_i + c$, $\bar{y} = \bar{x} + c$. Pentru această serie, dispersia va fi:

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum [(x_i + c) - (\bar{x} + c)]^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n} = s_x^2 \qquad (3.99.)$$

- Fie o distribuţie de frecvenţe, de forma $\begin{pmatrix} x_i \\ n_i \end{pmatrix}_{i=\overline{1,r}}$, cu media $\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$ şi dispersia $s_x^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i}$, şi se măreşte fiecare termen cu o constantă c.

Obţinem atunci o nouă distribuţie de forma: $(y_i)_{i=\overline{1,r}}$, unde $y_i = x_i + c$, $\bar{y} = \bar{x} + c$. Pentru această serie, dispersia va fi:

Scanned with OKEN Scanner

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2 n_i}{\sum n_i} = \frac{\sum [(x_i + c) - (\bar{x} + c)]^2 n_i}{\sum n_i} = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i} = s_x^2 \quad (3.100.)$$

3. Dacă într-o serie (simplă sau de frecvențe), pentru care s-a calculat media $\bar{x}$ și dispersia s_x^2, fiecare termen se mărește sau se micșorează de un același număr de ori (m), atunci dispersia seriei se va modifica și ea în același sens de m^2 ori, față de dispersia seriei inițiale.

- Fie o serie simplă, de forma $(x_i)_{i=\overline{1,n}}$, cu media $\bar{x} = \frac{\sum x_i}{n}$ și dispersia $s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$. Dacă fiecare termen se multiplică de m ori, atunci noua serie, va fi de forma: $(y_i)_{i=\overline{1,n}}$, unde $y_i = m \cdot x_i$, cu media:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum m \cdot x_i}{n} = m \cdot \frac{\sum x_i}{n} = m \cdot \bar{x} \quad (3.101.)$$

și dispersia:

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum (m \cdot x_i - m \cdot \bar{x})^2}{n} = \frac{m^2 \cdot \sum (x_i - \bar{x})^2}{n} = m^2 \cdot s_x^2 \quad (3.102.)$$

- Fie o distribuție de frecvențe, de forma $\begin{pmatrix} x_i \\ n_i \end{pmatrix}_{i=\overline{1,r}}$, cu media $\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$ și dispersia $s_x^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i}$. Dacă fiecare termen se multiplică de m ori, atunci noua serie, va fi de forma: $\begin{pmatrix} y_i \\ n_i \end{pmatrix}_{i=\overline{1,r}}$, unde $y_i = m \cdot x_i$, cu media:

$$\bar{y} = \frac{\sum y_i n_i}{\sum n_i} = \frac{\sum m \cdot x_i n_i}{\sum n_i} = m \cdot \frac{\sum x_i n_i}{\sum n_i} = m \cdot \bar{x} \quad (3.103.)$$

și dispersia:

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2 n_i}{\sum n_i} = \frac{\sum (m \cdot x_i - m \cdot \bar{x})^2 n_i}{\sum n_i} = \frac{m^2 \cdot \sum (x_i - \bar{x})^2 n_i}{\sum n_i} = m^2 \cdot s_x^2 \quad (3.104.)$$

Scanned with OKEN Scanner

4. Dispersia se poate calcula şi cu ajutorul formulei de calcul simplificat, ce provine dintr-o îmbinare a proprietăţilor (2) şi (3):

- pentru serii simple:

$$s^2 = \frac{\sum\left(\frac{x_i - a}{h}\right)^2}{n} \cdot h^2 - \left(\bar{x} - a\right)^2 \qquad (3.105.)$$

- pentru serii de frecvenţe absolute:

$$s^2 = \frac{\sum\left(\frac{x_i - a}{h}\right)^2 n_i}{\sum n_i} \cdot h^2 - \left(\bar{x} - a\right)^2 \qquad (3.106.)$$

- pentru serii de frecvenţe relative:

$$s^2 = \frac{\sum\left(\frac{x_i - a}{h}\right)^2 n_i^*}{\sum n_i^*} \cdot h^2 - \left(\bar{x} - a\right)^2 \qquad (3.107.)$$

Demonstrarea formulelor se face imediat, prin desfacerea parantezelor în expresiile din dreapta egalităţilor.

Aşa cum am arătat şi în cazul mediei, formula de calcul simplificat a dispersiei se utilizează în cazul distribuţiilor de frecvenţe pe intervale de variaţie, alegând a centrul intervalului cu frecvenţa maximă sau centrul intervalului din mijlocul distribuţiei, iar h mărimea intervalului.

Exemplul 3.43. Pentru datele din Tabelul 3.3., se alege $a = 10$ şi $h = 4$. Calculele intermediare se regăsesc în Tabelul 3.14., coloanele 3-5.

Tabelul 3.14.

Nr. de ani de existenţă	Nr. de agenţi economici	x_i	$\frac{x_i - a}{h}$	$\left(\frac{x_i - a}{h}\right)^2$	$\left(\frac{x_i - a}{h}\right)^2 n_i$
0	1	2	3	4	5
0-4	17	2	–2	4	68
4-8	29	6	–1	1	29
8-12	80	10	0	0	0
12-16	37	14	1	1	37
16-20	27	18	2	4	108
20-24	10	22	3	9	90
Total	200	–	–	–	332

$$s^2 = \frac{332}{200} \cdot 16 - (11{,}16 - 10)^2 = 26{,}56 - 1{,}3456 = 25{,}21$$

Scanned with OKEN Scanner

5. Dacă într-o distribuție de frecvențe toate frecvențele se modifică (în sensul multiplicării sau simplificării) de același număr k de ori, atunci dispersia seriei rămâne neschimbată.

Fie o distribuție de frecvențe, de forma $\begin{pmatrix} x_i \\ n_i \end{pmatrix}_{i=\overline{1,r}}$, cu media $\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$ și dispersia $s^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i}$. Dacă toate frecvențele se simplifică de k ori, distribuția devine:

$$\begin{pmatrix} x_i \\ n_i' \end{pmatrix}_{i=\overline{1,r}}, \text{ cu } n_i' = \frac{n_i}{k} \tag{3.108.}$$

Media noii distribuții va fi $\bar{x}' = \bar{x}$ (conform proprietății mediei), iar dispersia:

$$s'^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i'}{\sum_{i=1}^{r} n_i'} = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 \frac{n_i}{k}}{\sum_{i=1}^{r} \frac{n_i}{k}} = \frac{\frac{1}{k} \sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{\frac{1}{k} \sum_{i=1}^{r} n_i} = s^2 \tag{3.109.}$$

6. Dispersia termenilor unei serii față de nivelul mediu este mai mică decât dispersia termenilor seriei față de orice valoare (c) cu pătratul diferenței dintre medie și acea valoare c.

Fie o serie simplă, de forma $(x_i)_{i=\overline{1,n}}$, cu media $\bar{x} = \frac{\sum x_i}{n}$ și dispersia $s_{/x}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$. Dispersia termenilor seriei față de o valoare c se scrie:

$$s_{/c}^2 = \frac{\sum (x_i - c)^2}{n} = \frac{\sum (x_i - \bar{x} + \bar{x} - c)^2}{n} = \frac{\sum \left[(x_i - \bar{x}) + (\bar{x} - c) \right]^2}{n} =$$

$$= \frac{\sum \left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2 \right]}{n} =$$

$$= \frac{\sum (x_i - \bar{x})^2}{n} + 2(\bar{x} - c) \cdot \frac{2 \cdot \sum (x_i - \bar{x})}{n} + \frac{n \cdot (\bar{x} - c)^2}{n} = s_{/x}^2 + (\bar{x} - c)^2 \tag{3.110.}$$

Scanned with OKEN Scanner

Observaţie: Dacă volumul colectivităţii studiate, de la care am înregistrat datele, este redus (adică sub 30 de unităţi), atunci dispersia variabilei pentru acest eşantion se determină raportând suma pătratelor abaterilor termenilor seriei faţă de medie, la $(n-1)$ în loc de (n), conform relaţiei:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{(pentru serii simple)} \qquad (3.111.)$$

sau:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i - 1} \quad \text{(pentru serii de frecvenţe)} \qquad (3.112.)$$

Numai în acest fel dispersia la nivelul eşantionului va estima suficient de corect dispersia la nivelul colectivităţii totale. Problema va fi detaliată în capitolul 4.

Dacă vrem să determinăm dispersia pentru o distribuţie de frecvenţe pe intervale de variaţie, valorile x_i vor fi înlocuite cu centrele de interval, făcându-se presupunerea că unităţile sunt normal distribuite în interiorul intervalelor. Prin această înlocuire se face o aproximare, care va modifica valoarea dispersiei astfel calculate faţă de cea reală. Pentru ca valoarea dispersiei să nu fie prea mult alterată din cauza efectului de grupare Sheppard propune corectarea ei cu a 12-a parte a pătratului mărimii intervalului de grupare, adică:

$$s^2_{corectat} = s^2 - \frac{h^2}{12} \qquad (3.113.)$$

Această corecţie trebuie aplicată în următoarele condiţii:

- distribuţia de frecvenţe se referă la date continue;
- datele sunt grupate pe intervale de mărime egală;
- distribuţia are forma „clopotului lui Gauss" (a distribuţiei normale, în care cele mai multe valori sunt de mărime medie şi frecvenţele scad spre capătul cu valori mici sau mari al distribuţiei).

Exemplul 3.44. Pentru dispersia calculată în exemplul (3.43.), aplicăm corecţia lui Sheppard pentru a elimina efectul de grupare:

$$s^2_{corectat} = s^2 - \frac{h^2}{12} = 25{,}21 - \frac{16}{12} = 23{,}88$$

Dispersia pentru o caracteristică alternativă, pentru care datele sunt sistematizate ca în Tabelul 3.8., va fi dată de:

$$s^2_{alt} = \frac{(1-p)^2 m + (0-p)^2 (n-m)}{n} = (1-p)^2 \frac{m}{n} + p^2 \frac{n-m}{n} =$$

$$= (1-p)^2 \cdot p + p^2 (1-p) = p(1-p) = \frac{m}{n}\left(1 - \frac{m}{n}\right) \qquad (3.114.)$$

Scanned with OKEN Scanner

unde p este media variabilei alternative, iar m este numărul unităților statistice la care întâlnim varianta „favorabilă" a caracteristicii.

Exemplul 3.45. Pentru problema din exemplul (3.25.), dispersia vizitatorilor la târgul de turism care în următoarea lună vor să apeleze la serviciile oferite de agenție, va fi:

$$s_{alt}^2 = p(1-p) = 0{,}25 \cdot 0{,}75 = 0{,}1875$$

III. ***Abaterea medie pătratică (sau abatere standard, sau deviație standard)*** media pătratică simplă sau ponderată a abaterilor individuale absolute față de medie:

- pentru o serie simplă:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad (3.115.)$$

- pentru o serie de frecvențe absolute:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i}} \qquad (3.116.)$$

- pentru o serie de frecvențe relative:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i^*}{\sum_{i=1}^{r} n_i^*}} \qquad (3.117.)$$

Exemplul 3.46. Pentru seria simplă din exemplul (3.34.), abaterea medie pătratică va fi:

$$s = \sqrt{s^2} = \sqrt{3{,}6} = 1{,}897 \text{ u.m.}$$

Un termen al seriei se abate în medie cu 1,897 u.m. de la valoarea medie.

Exemplul 3.47. Pentru distribuția de frecvențe din Tabelul 3.3., deviația standard este:

$$s = \sqrt{s^2} = \sqrt{25{,}21} = 5{,}02 \text{ u.m.}$$

Deviația standard se măsoară în unități de măsură concrete, aceeași unitate ca și caracteristica pentru care s-a calculat. De aceea, ca și abaterea medie liniară,

Scanned with OKEN Scanner

ea nu poate fi utilizată pentru compararea variației unor serii care se referă la variabile diferite, ale căror unități de măsură sunt diferite.

Deviația standard, fiind calculată pe baza unei medii pătratice, este, de regulă, mai mare decât abaterea medie liniară, ce reprezintă o medie aritmetică. Evident, acest lucru se întâmplă numai dacă ambii indicatori sunt calculați pentru aceeași serie:

$$s > \bar{d} \qquad (3.118.)$$

Pentru o distribuție cu tendință de normalitate, între abaterea medie liniară și cea pătratică există următoarea relație:

$$\bar{d} \cong \frac{4}{5} s \qquad (3.119.)$$

O regulă empirică de determinare a unei valori estimative a deviației standard într-o distribuție cu tendință de normalitate este aceea conform căreia raportul dintre amplitudine și abatere medie pătratică este rareori mai mic decât 2 și mai mare decât 6. Altfel spus, dacă împărțim amplitudinea la 6 obținem o aproximare (o estimare) a abaterii medii pătratice. Dacă valoarea obținută din calcul a abaterii medii pătratice într-o distribuție normală este mult diferită de valoarea determinată conform regulii empirice menționate, atunci este foarte probabil că sunt greșit efectuate calculele. Alături de medie, abaterea medie pătratică este cel de-al doilea indicator care definește, care descrie, practic, un model de distribuție de frecvențe.

În majoritatea cazurilor, în descrierea unor concepte statistice și a unor distribuții empirice, ne raportăm la distribuția normală, ca model teoretic, în care majoritatea observațiilor sunt concentrate în zona de centru a distribuției. Pe măsură ce ne depărtăm de centrul distribuției, frecvențele observațiilor descresc de o parte și de alta a centrului, către valorile extreme (către ambele capete ale distribuției) [Figura 3.10.)].

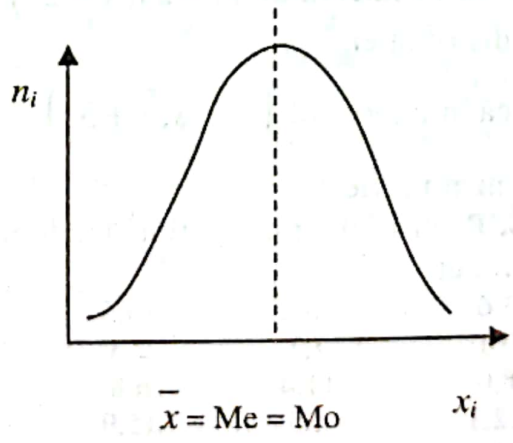


Figura 3.10. Repartiția normală

Scanned with OKEN Scanner

Distribuția normală este descrisă de funcția:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-(x-\mu)^2/2\sigma^2} \qquad (3.120.)$$

unde:

$f(x)$ este frecvența relativă a unei anumite valori/variante;
x este valoarea/varianta observată;
μ este media caracteristicii la nivelul populației totale;
σ^2 este dispersia caracteristicii la nivelul populației totale.

Conform regulii empirice, care se aplică în cazul variabilelor distribuite normal sau asimptotic normal, există anumite proporții ale curbei între diferitele părți ale distribuției. Astfel, raportul dintre aria cuprinsă între medie și:

- $\pm 1{,}0 \cdot s$ este de aproximativ 0,68 (68%);
- $\pm 2{,}0 \cdot s$ este de aproximativ 0,95 (95%);
- $\pm 3{,}0 \cdot s$ este de aproximativ 0,997 (99,7%);

Altfel exprimat, 68% dintre valorile caracteristicii urmărite se găsesc în intervalul: $(\bar{x}-s;\bar{x}+s)$, 95% dintre observații se regăsesc în intervalul $(\bar{x}-2s;\bar{x}+2s)$, iar 99,7% dintre observații sunt în intervalul $(\bar{x}-3s;\bar{x}+3s)$.

Pentru orice distribuție de date, putem însă aplica teorema lui Cebîșev, care spune că: în intervalul aflat la o distanță de k deviații standard față de medie $(\bar{x}-k\cdot s;\bar{x}+k\cdot s)$ se găsesc întotdeauna cel puțin $\left(1-\frac{1}{k^2}\right)\%$ dintre termenii seriei, unde k este orice număr pozitiv mai mare decât 1. Altfel spus, pentru:

– $k = 2$ rezultă că în intervalul $(\bar{x}-2s;\bar{x}+2s)$ se găsesc cel puțin 75% dintre termenii distribuției;

– $k = 3$ rezultă că în intervalul $(\bar{x}-3s;\bar{x}+3s)$ se găsesc cel puțin $\frac{8}{9}$ (adică 89%) dintre termenii seriei.

Exemplul 3.48. Pentru 30 de convorbiri telefonice de lungă-distanță s-au înregistrat duratele (în minute):

11,8	3,6	16,6	13,5	4,8	8,3
8,9	9,1	7,7	2,3	12,1	6,1
10,2	8,0	11,4	6,8	9,6	19,5
15,3	12,3	8,5	15,9	18,7	11,7
6,2	11,2	10,4	7,2	5,5	14,5

Scanned with OKEN Scanner

Să se determine ponderea cazurilor care se găsesc în intervalul:

$(\bar{x}-s;\bar{x}+s)$

$(\bar{x}-2s;\bar{x}+2s)$

$(\bar{x}-3s;\bar{x}+3s)$

şi să se stabilească dacă valorile găsite corespund regulii empirice a lui Cebîşev.

Rezolvare:

Valoarea medie, calculată pe baza datelor iniţiale este:

$$\bar{x}=\frac{\sum x_i}{n}=10{,}26 \text{ minute}$$

iar abaterea medie pătratică:

$$s=\sqrt{\frac{\sum\left(x_i-\bar{x}\right)^2}{n}}=4{,}29 \text{ minute}$$

$(\bar{x}-s;\bar{x}+s)=(10{,}26\text{-}4{,}29;\ \ 10{,}26+4{,}29)=(5{,}97;\ \ 14{,}55)$

$(\bar{x}-2s;\bar{x}+2s)=(1{,}68;\ \ 18{,}84)$

Tabelul 3.15.

Interval	Ponderea convorbirilor telefonice pentru distribuţia dată	Regula empirică
$(\bar{x}-s;\bar{x}+s)=(5{,}97;\ \ 14{,}55)$	21/30 = 0,7 (70%)	68%
$(\bar{x}-2s;\bar{x}+2s)=(1{,}68;\ \ 18{,}84)$	29/30 = 0,967 (96,7%)	95%

Se observă că procentele calculate pe baza datelor iniţiale sunt foarte apropiate de cele corespunzătoare regulii empirice. În ultimul interval, $(x-3s;x+3s)$ se găsesc practic toate convorbirile telefonice.

IV. Dacă abaterea medie pătratică şi abaterea medie liniară măsoară sub formă absolută dispersarea valorilor, ***coeficientul de variaţie*** este indicatorul sintetic al variaţiei care măsoară în mod relativ şi sintetic gradul de împrăştiere a valorilor faţă de tendinţa centrală a seriei. El se determină ca raport între deviaţia standard şi media aritmetică; se poate exprima şi în procente, important este faptul că nemaidepinzând de unitatea de măsură a caracteristicii urmărite, poate fi folosit pentru a compara omogenitatea sau, dimpotrivă, eterogenitatea a două sau mai multe serii, care se referă la variabile diferite.

$$v=\frac{s}{\bar{x}}\cdot 100 \qquad (3.121.)$$

sau

$$v'=\frac{\bar{d}}{\bar{x}}\cdot 100 \qquad (3.122.)$$

Scanned with OKEN Scanner

O valoare mai mică a coeficientului de variație indică un grad de variație mai mic al seriei statistice și implicit, un nivel de omogenitate crescut. Astfel, dacă valoarea coeficientului de variație este mai mică sau cel mult egală cu 30-35%, atunci seria este omogenă și media este reprezentativă pentru valorile din care s-a calculat. Dacă, dimpotrivă, valoarea coeficientului de variație este de peste 65-70%, seria este eterogenă, media calculată își pierde semnificația și nu mai este reprezentativă.

Exemplul 3.49. Pentru seriile statistice $X = \{7,8,9\}$ și $Y = \{2,8,14\}$ se observă cu ochiul liber că, deși ambele serii au aceeași medie, a doua serie prezintă o împrăștiere mai mare decât prima serie.

Într-adevăr, $\bar{x} = \bar{y} = 8$

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{(7-8)^2 + (8-8)^2 + (9-8)^2}{3}} = \sqrt{0,67} = 0,82$$

$$v_x = \frac{s_x}{\bar{x}} \cdot 100 = \frac{0,82}{8} \cdot 100 = 10,23\%$$

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n}} = \sqrt{\frac{(2-8)^2 + (8-8)^2 + (14-8)^2}{3}} = \sqrt{24} = 4,899$$

$$v_y = \frac{s_y}{\bar{y}} \cdot 100 = \frac{4,899}{8} \cdot 100 = 61,24\%$$

Cum $v_x < v_y$ rezultă că seria X este mai omogenă decât Y.

Exemplul 3.50. Pentru seria simplă din exemplul (3.34.), coeficientul de variație este:

$$v = \frac{s}{\bar{x}} \cdot 100 = \frac{1,897}{9} \cdot 100 = 21,08\% < 35\%$$

Rezultă că seria este omogenă, media este reprezentativă.

Exemplul 3.51. Pentru distribuția de frecvențe din Tabelul 3.3., coeficientul de variație este:

$$v = \frac{s}{\bar{x}} \cdot 100 = \frac{5,02}{11,16} \cdot 100 = 45\% > 35\%$$

Rezultă că seria nu este omogenă, media nu este reprezentativă.

3.3.3. Măsurarea dispersării valorilor unei caracteristici într-o populație statistică structurată pe grupe

Am văzut că fenomenele de masă, studiate de statistică, sunt rezultatul acțiunii unui număr mare de factori de influență. Cu cât fenomenele sunt deter-

Scanned with OKEN Scanner

minate de un număr mai mare de factori, cu atât prezintă un nivel de complexitate mai ridicat şi, implicit, un grad mai mare de variaţie a valorilor caracteristicilor care le definesc. Deoarece populaţia statistică, în ansamblul ei, prezintă un grad de împrăştiere a valorilor ridicat, ea trebuie structurată pe grupe de unităţi, în funcţie de modul de variaţie a factorilor determinanţi. Este de dorit obţinerea unor grupe omogene (cu un grad de variaţie cât mai redus în cadrul lor), dar care se comportă diferit unele de altele. Este, deci, de aşteptat, ca variabila să prezinte şi o variaţie de la o grupă la alta, pe lângă cea din interiorul grupelor. Variaţia caracteristicii în cadrul fiecărei grupe este pusă pe seama factorilor aleatorii, în timp ce variaţia dintre grupe este explicată prin acţiunea factorului în funcţie de care s-a efectuat gruparea (de factorul principal de grupare).

Să presupunem că de la unităţile unei populaţii statistice s-au înregistrat valorile a două caracteristici X şi Y (unde X este un factor de influenţă al lui Y); colectivitatea a fost apoi împărţită în r grupe în funcţie de valorile variabilei cauzale X. Datele sistematizate se prezintă ca în Tabelul 3.16.

Distribuţia de frecvenţe bidimensională

Tabelul 3.16.

Variante sau centre de interval pt. variabila X	Variante sau centre de interval pt. variabila Y						Total unităţi	Medii de grupă ale var. Y	Dispersii de grupă ale var. Y
	y_1	y_2	...	y_j	...	y_p			
0	1	2	3	4	5	6	7	8	9
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$	$\bar{y}_1$	s^2_{y1}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$	$\bar{y}_2$	s^2_{y2}
...	...	...	...	...	...	...	...	...	
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$	$\bar{y}_i$	s^2_{yi}
...	...	...	...	...	...	...	...	...	
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rp}	$n_{r.}$	$\bar{y}_r$	s^2_{yr}
Total unităţi	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.p}$	$n_{..}$	$\bar{y}$	s^2_y

Simbolurile utilizate în Tabelul 3.16., coloanele 0-7 sunt explicate în Tabelul 2.12. din capitolul 2.

La nivelul fiecărei grupe i a colectivităţii ($i = \overline{1, r}$), se poate calcula:

- o medie a valorilor variabilei Y înregistrate la unităţile din acea grupă:

Scanned with OKEN Scanner

$$\bar{y}_i = \frac{\sum_{j=1}^{p} y_j n_{ij}}{\sum_{j=1}^{p} n_{ij}} = \frac{\sum_{j=1}^{p} y_j n_{ij}}{n_i} \qquad (3.123.)$$

(Tabelul 3.16., coloana 8).

- o dispersie a valorilor variabilei Y (y_j) înregistrate la unităţile grupei i, în jurul mediei de grupă $\bar{y}_i$

$$s_{yi}^2 = \frac{\sum_{j=1}^{p} (y_j - \bar{y}_i)^2 n_{ij}}{\sum_{j=1}^{p} n_{ij}} \qquad (3.124.)$$

(Tabelul 3.16., coloana 9).

Mediile şi dispersiile de grupă, calculate prin relaţiile (3.123.) şi (3.124.) se mai numesc ***medii şi dispersii parţiale sau condiţionate*** de factorul de grupare X.

Pe baza mediilor şi dispersiilor condiţionate, se determină media şi dispersia totală a lui Y, la nivelul întregii colectivităţi.

- media valorilor variabilei Y pe ansamblul populaţiei studiate se poate determina în două moduri:
  - fie ca medie aritmetică ponderată a variantelor sau valorilor variabilei Y (y_j), rolul de ponderi fiind jucat de frecvenţele de apariţie ale acestor variante sau valori ($n_{.j}$), pe ansamblul colectivităţii:

$$\bar{y} = \frac{\sum_{j=1}^{p} y_j n_{.j}}{\sum_{j=1}^{p} n_{.j}} = \frac{\sum_{j=1}^{p} y_j n_{.j}}{n} \qquad (3.125.)$$

  - fie ca medie aritmetică ponderată a mediilor condiţionate ale variabilei Y ($\bar{y}_i$), rolul de ponderi fiind jucat de volumul sau frecvenţa grupelor (n_i):

$$\bar{y} = \frac{\sum_{i=1}^{r} \bar{y}_i n_i}{\sum_{i=1}^{r} n_i} \qquad (3.126.)$$

Scanned with OKEN Scanner

Dacă toate grupele conţin acelaşi număr de unităţi (au volum egal), atunci media totală a valorilor variabilei Y se calculează ca medie aritmetică simplă:

$$\bar{y} = \frac{\sum_{i=1}^{r} \bar{y}_i}{r} \qquad (3.127.)$$

Spre deosebire de media totală a variabilei Y, care se poate determina, aşa cum am văzut, ca medie a mediilor parţiale (condiţionate de factorul de grupare X) ale variabilei Y, dispersia totală a acestei variabile se poate calcula pe seama dispersiilor caracteristicii Y la nivelul grupelor, dar printr-o relaţie ceva mai complexă.

Astfel, măsurarea variabilităţii totale a caracteristicii Y prin intermediul dispersiei totale porneşte de la principiul identificării principalelor surse ale variaţiei variabilei rezultative Y. Dispersarea valorilor acestei caracteristici se produce, pe de o parte, sub influenţa factorului principal de grupare Y (factor determinant), iar pe de altă parte sub influenţa acţiunii factorilor aleatorii (adică toţi ceilalţi factori care îl influenţează pe Y, mai puţin Y). Prima categorie de factori (X) induce o variaţie a lui Y de la o grupă la alta (sau între grupe), în timp ce a doua categorie de factori induce o variaţie a valorilor lui Y în interiorul fiecărei grupe după X.

- ***Variabilitatea din cadrul grupelor*** se determină cu ajutorul unui indicator numit ***media dispersiilor grupelor***, calculat ca medie aritmetică simplă sau ponderată a dispersiilor grupelor (condiţionate de factorul de grupare X):

$$\bar{s}^2 = \frac{\sum_{i=1}^{r} s_{yi}^2 n_i}{\sum_{i=1}^{r} n_i} \quad \text{(dacă volumele grupelor sunt diferite)} \qquad (3.128.)$$

$$\bar{s}^2 = \frac{\sum_{i=1}^{r} s_{yi}^2}{r} \quad \text{(dacă volumele grupelor sunt egale)} \qquad (3.129.)$$

- ***Variabilitatea caracteristicii Y prezentată de la o grupă la alta*** (sau variabilitatea dintre grupe) se măsoară cu ajutorul indicatorului ***dispersia dintre grupe***, calculată ca medie aritmetică (simplă sau ponderată) a abaterilor mediilor de grupă (parţiale, condiţionate) faţă de media totală a variabilei Y.

$$\delta^2 = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{y})^2 n_i}{\sum_{i=1}^{r} n_i} \qquad (3.130.)$$

Scanned with OKEN Scanner

La fel ca în cazul mediei dispersiilor grupelor, dacă grupele au volume egale, media aritmetică ponderată se transformă în medie aritmetică simplă:

$$\delta^2 = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{\bar{y}})^2}{r} \qquad (3.131.)$$

- ***Variabilitatea totală a caracteristicii Y*** măsurată prin ***dispersia totală*** a acesteia se determină adunând dispersia dintre grupe şi media dispersiilor grupelor, într-o relaţie denumită ***Regula adunării dispersiilor***:

$$s_y^2 = \bar{s}^2 + \delta^2 \qquad (3.132.)$$

Dispersia totală a variabilei Y se mai poate determina şi cu ajutorul distribuţiei marginale $\begin{pmatrix} y_j \\ n_{.j} \end{pmatrix}_{j=\overline{1,p}}$, ca medie aritmetică ponderată a abaterilor valorilor individuale (y_j) faţă de media totală ($\bar{\bar{y}}$):

$$s_y^2 = \frac{\sum_{j=1}^{p} (y_j - \bar{\bar{y}})^2 n_{.j}}{\sum_{j=1}^{p} n_{.j}} \qquad (3.133.)$$

Toate formulele descrise în această secţiune se menţin şi în cazul utilizării frecvenţelor relative.

Din relaţia (3.132.) se observă că, dacă dispersia dintre grupe are o valoare superioară mediei dispersiilor parţiale, atunci înseamnă că şi factorul de grupare X a avut o influenţă mai mare asupra variabilei rezultative Y, decât factorii aleatorii. Dacă, dimpotrivă, media dispersiilor condiţionate este mai mare decât dispersia dintre grupe, atunci se poate spune că variabila Y a fost influenţată, în variaţia sa, în mai mare măsură de factorii aleatorii, decât de factorul determinant X. Pentru a determina în ce grad influenţează factorul de grupare X variaţia variabilei Y, calculăm coeficientul (gradul, raportul) de determinaţie, ca raport între dispersia dintre grupe şi dispersia totală:

$$R^2 = \frac{\delta^2}{s_y^2} \qquad (3.134.)$$

Scanned with OKEN Scanner

Raportul de determinație se poate exprima și procentual:

$$R^2_{\%} = \frac{\delta^2}{s_y^2} \cdot 100 \qquad (3.135.)$$

Pentru a determina în ce măsură influențează ceilalți factori variația variabilei Y, se calculează coeficientul (gradul, raportul) de nedeterminație, ca raport între media dispersiilor parțiale și dispersia totală a variabilei Y:

$$K^2 = \frac{\bar{s}^2}{s_y^2} \qquad (3.136.)$$

Raportul de nedeterminație se poate exprima și procentual:

$$K^2_{\%} = \frac{\bar{s}^2}{s_y^2} \cdot 100 \qquad (3.137.)$$

Între cele două rapoarte există relația:

$$R^2 + K^2 = 1 \qquad (3.138.)$$

sau

$$R^2_{\%} + K^2_{\%} = 100 \qquad (3.139.)$$

$$K^2 = 1 - R^2 \qquad (3.140.)$$

sau

$$K^2_{\%} = 100 - R^2_{\%} \qquad (3.141.)$$

Dacă $R^2 > K^2$ atunci spunem că factorul de grupare X este determinant, iar dacă $K^2 > R^2$, atunci factorii aleatorii sunt cei care influențează preponderent variația variabilei Y.

Exemplul 3.52. În cele 20 de farmacii existente în sectorul A al unei localități, prețul mediu de vânzare al unui medicament este de 13,5 RON, cu o dispersie de 5,75; în același timp, în cele 30 de farmacii din sectorul B al aceleiași localități, același medicament se vinde la un preț mediu de 8,30 RON, cu o dispersie a prețurilor de 2,84. Se pune problema determinării influenței pe care o are sectorul asupra variației prețurilor la medicamentul studiat.

Pentru aceasta, vom face următoarele notații:

- numărul de farmacii din sectorul A: $n_A = 20$;
- prețul mediu de vânzare al medicamentului în sectorul A: $\bar{y}_A = 12{,}5\,\text{RON}$;

Scanned with OKEN Scanner

- dispersia preţurilor medicamentului în farmaciile din sectorul A: $s_{yA}^2 = 5{,}75$;
- numărul de farmacii din sectorul B: $n_B = 30$;
- preţul mediu de vânzare al medicamentului în sectorul B: $\bar{y}_B = 8{,}30$ RON;
- dispersia preţurilor medicamentului în farmaciile din sectorul B: $s_{yA}^2 = 2{,}84$.

Preţul mediu cu care s-a vândut medicamentul pe ansamblul celor două sectoare a fost:

$$\bar{y} = \frac{\sum_{i=1}^{r} \bar{y}_i n_i}{\sum_{i=1}^{r} n_i} = \frac{\bar{y}_A n_A + \bar{y}_B n_B}{n_A + n_B} = \frac{13{,}5 \cdot 20 + 8{,}3 \cdot 30}{50} = 10{,}38 \text{ RON}$$

Media dispersiilor parţiale va fi:

$$\bar{s}^2 = \frac{\sum_{i=1}^{r} s_{yi}^2 n_i}{\sum_{i=1}^{r} n_i} = \frac{s_{yA}^2 n_A + s_{yB}^2 n_B}{n_A + n_B} = \frac{5{,}75 \cdot 20 + 2{,}84 \cdot 30}{50} = 4$$

Dispersia dintre grupe va fi:

$$\delta^2 = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{y})^2 n_i}{\sum_{i=1}^{r} n_i} = \frac{(\bar{y}_A - \bar{y})^2 n_A + (\bar{y}_B - \bar{y})^2 n_B}{n_A + n_B} =$$

$$= \frac{(13{,}5 - 10{,}38)^2 \cdot 20 + (8{,}30 - 10{,}38)^2 \cdot 30}{50} = 6{,}49$$

Dispersia totală a preţului medicamentului va fi:

$$s_y^2 = \bar{s}^2 + \delta^2 = 4{,}00 + 6{,}49 = 10{,}49$$

Gradul în care sectorul influenţează variaţia preţului este dat de:

$$R^2 = \frac{\delta^2}{s_y^2} = \frac{6{,}49}{10{,}49} = 0{,}6186 \text{ (adică 61, 86\%)}$$

Scanned with OKEN Scanner

3.4. Analiza formei distribuțiilor de frecvențe

În analiza formei distribuțiilor de frecvențe, se pleacă de la reprezentarea grafică a acestora (histograma sau poligonul frecvențelor). Distribuțiile de frecvențe pot avea o multitudine de forme, dar, cel mai adesea, ele sunt comparate cu distribuția normală. Unele dintre cele mai importante caracteristici ale unei distribuții sunt boltirea (excesul) și asimetria (oblicitatea).

3.4.1. Boltirea (excesul)

Boltirea (concentrarea) caracterizează măsura în care valorile sunt concentrate în jurul unei porțiuni a scalei de valori sau, dimpotrivă, sunt dispersate de-a lungul întregii scale. O distribuție poate fi în același timp de tip normal, dar să prezinte grade diferite de concentrare a variantelor în jurul unor anumite valori [Figura 3.11.)].

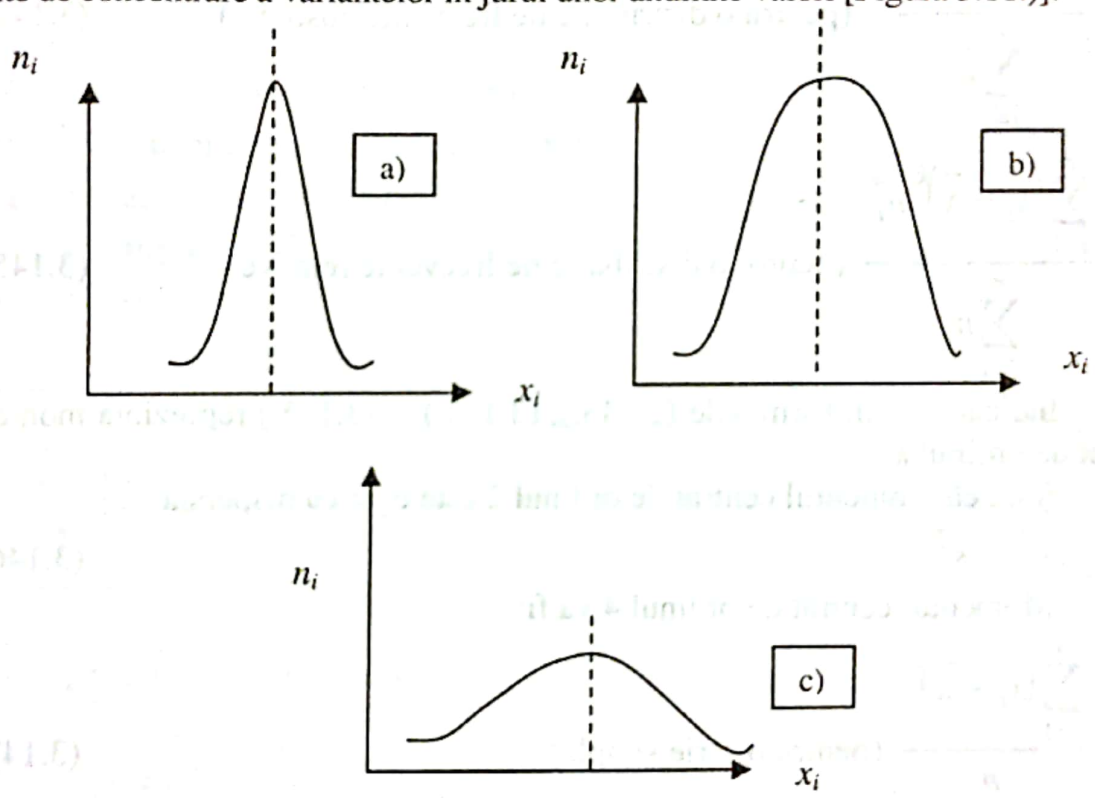


Figura 3.11. Reprezentarea grafică a unei distribuții normale cu grade diferite de concentrare a valorilor: a) leptocurtică; b) mezocurtică; c) platicurtică.

În Figura 3.11. a) este prezentată o distribuție leptocurtică, cu o formă ascuțită, țuguiată, în care există o aglomerare mai pronunțată a variantelor/valorilor în centrul distribuției; în Figura 3.11. c) este reprezentată o distribuție platicurtică, cu o formă mai aplatizată, în care variantele caracteristicii sunt dispersate de-a lungul scalei; distribuția din Figura 3.11. b) – mezocurtică prezintă o situație de

Scanned with OKEN Scanner

mijloc între curbele a) şi c). Măsurarea gradului de concentrare a valorilor unei repartiţii de frecvenţe se realizează cu ajutorul indicatorului:

$$\beta = \frac{c_4}{c_2^2} \tag{3.142.}$$

unde:

$$c_k = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^k}{n} \text{ (pentru o serie simplă)} \tag{3.143.}$$

$$c_k = \frac{\sum_{i=1}^{r} \left(x_i - \bar{x}\right)^k n_i}{\sum_{i=1}^{r} n_i} \text{ (pentru o distribuţie de frecvenţe absolute)} \tag{3.144.}$$

$$c_k = \frac{\sum_{i=1}^{r} \left(x_i - \bar{x}\right)^k n_i^*}{\sum_{i=1}^{r} n_i^*} \text{ (pentru o distribuţie de frecvenţe relative)} \tag{3.145.}$$

Indicatorii din formulele (3.143.), (3.144.) şi (3.145.) reprezintă momentul centrat de ordinul k.

Ştim că momentul centrat de ordinul 2 este egal cu dispersia:

$$c_2 = s^2 \tag{3.146.}$$

Momentul centrat de ordinul 4 va fi:

$$c_4 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^4}{n} \text{ (pentru o serie simplă)} \tag{3.147.}$$

$$c_4 = \frac{\sum_{i=1}^{r} \left(x_i - \bar{x}\right)^4 n_i}{\sum_{i=1}^{r} n_i} \text{ (pentru o distribuţie de frecvenţe absolute)} \tag{3.148.}$$

Scanned with OKEN Scanner

$$c_4 = \frac{\sum_{i=1}^{r}\left(x_i - \bar{x}\right)^4 n_i^*}{\sum_{i=1}^{r} n_i^*} \quad \text{(pentru o distribuție de frecvențe relative)} \qquad (3.149.)$$

Pentru distribuția normală indicatorul boltirii β are valoarea 3 ($\beta = 3$). Dacă $\beta < 3$ atunci distribuția este platicurtică (există o dispersare mai mare a valorilor de-a lungul scalei decât în distribuția normală și curba va fi mai turtită); dacă $\beta > 3$ atunci distribuția este leptocurtică (există o concentrare mai ridicată a variantelor în jurul unei porțiuni a scalei decât în distribuția normală și curba este mai ascuțită).

Un alt indicator de măsurare a gradului de boltire se calculează pe baza indicatorului precedent în forma:

$$\gamma = \beta - 3 \qquad (3.150.)$$

Dacă $\gamma = 0$ atunci distribuția este normală;

Dacă $\gamma < 0$ atunci distribuția este platicurtică;

Dacă $\gamma > 0$ atunci distribuția este leptocurtică.

Exemplul 3.53. Pentru seria din exemplul (3.34.), indicatorul boltirii se determină (Tabelul 3.17.):

Tabelul 3.17.

x_i	7	7	9	10	12	Total 45
$\left(x_i - \bar{x}\right)^2$	4	4	0	1	9	18
$\left(x_i - \bar{x}\right)^4$	16	16	0	1	81	114

$$c_2^2 = \left(s^2\right)^2 = 3{,}6^2 = 12{,}96$$

$$c_4 = \frac{\sum\left(x_i - \bar{x}\right)^4}{n} = \frac{114}{5} = 22{,}8$$

$$\beta = \frac{c_4}{c_2^2} = \frac{22{,}8}{12{,}96} = 1{,}76 < 3$$

Rezultă că distribuția este platicurtică.

Exemplul 3.54. Pentru distribuția de frecvențe din Tabelul 3.3. determinăm indicatorul de boltire. Calculele se regăsesc în Tabelul 3.18.

Scanned with OKEN Scanner

Tabelul 3.18.

Număr de ani de existență (ani)	Nr. de agenți economici (n_i)	x_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4 n_i$
0	1	2	3	4
0-4	17	2	83,91	119.688,80
4-8	29	6	26,63	20.562,14
8-12	80	10	1,35	145,32
12-16	37	14	8,07	2408,31
16-20	27	18	46,79	59.105,64
20-24	10	22	117,51	138.080,83
Total	200	–	–	339.991,04

$$c_2^2 = \left(s^2\right)^2 = 25{,}21^2 = 635{,}54$$

$$c_4 = \frac{\sum (x_i - \bar{x})^4 n_i}{\sum n_i} = \frac{339.991{,}04}{200} = 1699{,}96$$

$$\beta = \frac{c_4}{c_2^2} = \frac{1699{,}96}{635{,}54} = 2{,}67 < 3$$

Distribuția este platicurtică.

3.4.2. Asimetria (oblicitatea)

Oblicitatea unei distribuții se referă la faptul că valorile nu sunt repartizate simetric față de tendința centrală, manifestându-se concentrarea, aglomerarea majorității valorilor la unul din capetele distribuției (spre valorile mici sau spre cele mari ale scalei). În secțiunea 3.2.6. a acestui capitol am prezentat cele trei situații posibile de distribuții de frecvențe, sub aspectul asimetriei (distribuție pozitiv asimetrică, distribuție negativ asimetrică și distribuție perfect simetrică). Până acum am analizat asimetria (oblicitatea) unei distribuții doar din punctul de vedere al direcției, al sensului acesteia. Există însă posibilitatea determinării și a gradului, a nivelului, a magnitudinii asimetriei, cu ajutorul unor indicatori absoluți și relativi:

A. **Indicatorii absoluți pentru măsurarea asimetriei sunt:**

- *asimetria absolută*, calculată ca diferență între valoarea medie și modală:

$$As = \bar{x} - Mo \qquad (3.151.)$$

Uneori, acest indicator se mai scrie în forma echivalentă:

$$As' = 3\left(\bar{x} - Me\right) \qquad (3.152.)$$

Scanned with OKEN Scanner

- *asimetria absolută*, calculată pe baza cuartilelor:

$$As^q = Q_3 + Q_1 - 2Me \qquad (3.153.)$$

Indicatorii definiți cu relațiile (3.151.)-(3.153.) sunt exprimați în unitatea de măsură a caracteristicii pentru care s-au calculat. Cu cât valoarea lor este mai mare, cu atât asimetria distribuției este mai accentuată.

B. Indicatorii relativi pentru măsurarea asimetriei sunt:

- *Coeficientul de asimetrie Pearson*, determinat ca raport între indicatorii absoluți ai asimetriei și media aritmetică a distribuției:

$$Cas^p = \frac{\bar{x} - Mo}{s} \qquad (3.154.)$$

sau

$$Cas'^p = \frac{3(\bar{x} - Me)}{s} \qquad (3.155.)$$

Relația (3.155.) se folosește numai pentru seriile cu tendință de normalitate.

Indicatorul Cas^p ia valori în intervalul $[-1,+1]$.

Dacă $Cas^p = 0 \Rightarrow \bar{x} = Me = Mo$, seria este perfect simetrică.

Dacă $Cas^p > 0 \Rightarrow \bar{x} > Me > Mo$, seria este pozitiv asimetrică, în ea predominând valorile mari.

Dacă $Cas^p < 0 \Rightarrow \bar{x} < Me < Mo$, seria este negativ asimetrică, în ea predominând valorile mici.

Dacă Cas^p ia valori apropiate de ±1, distribuția are asimetrie pronunțată (pozitivă sau negativă.

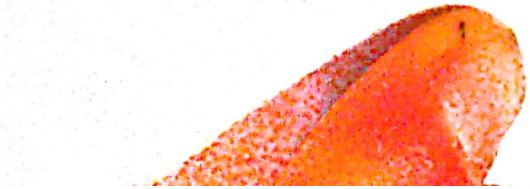
Dacă Cas^p ia valori apropiate de 0, distribuția are asimetrie ușoară (pozitivă sau negativă).

Cu cât indicatorul are valori mai apropiate de zero, în valoare absolută, seria este mai puțin asimetrică și, cu cât el are valori mai apropiate de capetele intervalului (de –1 sau de +1), cu atât distribuția este mai pronunțat asimetrică.

- *coeficientul de asimetrie Pearson* (exprimat cu ajutorul cuartilelor):

$$Cas_q^p = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1} \qquad (3.156.)$$

Indicatorul ia întotdeauna valori în intervalul $[-1,+1]$. Dacă el ia valori apropiate de 0,1, atunci seria este moderat asimetrică, iar dacă ia valori mai mari de 0,3, atunci seria este pronunțat asimetrică.



Scanned with OKEN Scanner

- *coeficientul de asimetrie Fisher*

$$Cas^F = \frac{c_3^2}{c_2^3} \qquad (3.157.)$$

unde:

c_3 este momentul centrat de ordinul 3;

c_2 este momentul centrat de ordinul 2 (dispersia).

Momentul centrat de ordinul 3 va fi definit conform relațiilor:

$$c_3 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{n} \text{ (pentru o serie simplă)} \qquad (3.158.)$$

$$c_3 = \frac{\sum_{i=1}^{r}\left(x_i - \bar{x}\right)^3 n_i}{\sum_{i=1}^{r} n_i} \text{ (pentru o distribuție de frecvențe absolute)} \qquad (3.159.)$$

$$c_3 = \frac{\sum_{i=1}^{r}\left(x_i - \bar{x}\right)^3 n_i^*}{\sum_{i=1}^{r} n_i^*} \text{ (pentru o distribuție de frecvențe relative)} \qquad (3.160.)$$

Dacă $Cas^F = 0$, atunci seria este simetrică.

Cu cât Cas^F diferă mai mult de zero, cu atât seria este mai pronunțat asimetrică.

Exemplul 3.55. Pentru distribuția de frecvențe din Tabelul 3.3., s-a obținut: $\bar{x} = 11{,}16$ ani; $Mo = 10{,}17$ ani; $s = 5{,}02$ ani. Coeficientul de asimetrie Pearson va fi:

$$Cas^P = \frac{\bar{x} - Mo}{s} = \frac{11{,}16 - 10{,}17}{5{,}02} = 0{,}197$$

Distribuția de frecvențe are o asimetrie pozitivă destul de ușoară.

Într-o distribuție de frecvențe prezintă interes și analiza modalității de distribuire a frecvențelor de apariție a variantelor. Astfel, frecvențele se pot repartiza uniform, neexistând deci puncte de concentrare în jurul unor anumite valori, sau, dimpotrivă, pot exista unul sau mai multe puncte de concentrare în jurul unor variante ale caracteristicii. Reprezentarea grafică se realizează cu ajutorul Curbei lui Lorentz (numită și curbă de concentrare), care se construiește alegând pe axa absciselor frecvențele relative cumulate ale seriei, iar pe axa ordonatelor – greutățile specifice cumulate ale valorilor centralizate, agregate ale caracteristicii.

Scanned with OKEN Scanner

Prin unirea tuturor punctelor ce au drept coordonate valorile cumulate ale variabilelor, de pe cele două axe, se obține curba de concentrare. Dacă această curbă este apropiată ca formă de o linie dreaptă (este foarte puțin arcuită), atunci frecvențele se distribuie uniform, apărând practic o lipsă a concentrării. Dar cu cât această curbă este mai concavă, cu atât există o concentrare mai mare, mai accentuată a frecvențelor distribuției.

Exemplul 3.56. Pentru distribuția de frecvențe din Tabelul 3.3. se cunoaște și numărul angajaților pe fiecare categorie de agenți economici după vechime. Datele se regăsesc în Tabelul 3.19.

Tabelul 3.19.

Număr de ani de existență (ani)	Nr. de agenți economici (n_i)	Frecvențe relative (n_i^*) (%)	Frecvențe relative cumulate (Fc_i^*)	Număr total de angajați	Greutatea specifică a angajaților (ponderi) (g_i) (%)	Greutățile specifice cumulate (g_i^{cum})
0	1	2	3	4	5	6
0-4	17	8,5	8,5	850	9,39	9,39
4-8	29	14,5	23,0	1223	13,52	22,91
8-12	80	40,0	63,0	3090	34,15	57,06
12-16	37	18,5	81,5	1722	19,03	76,09
16-20	27	13,5	95,0	1520	16,80	92,89
20-24	10	5,0	100,0	644	7,11	100,0
Total	200	100,0	–	9049	100,0	–

Curba Lorentz este trasată în Figura 3.12. și arată, datorită formei sale foarte puțin arcuite, o distribuție aproximativ uniformă a unităților statistice.

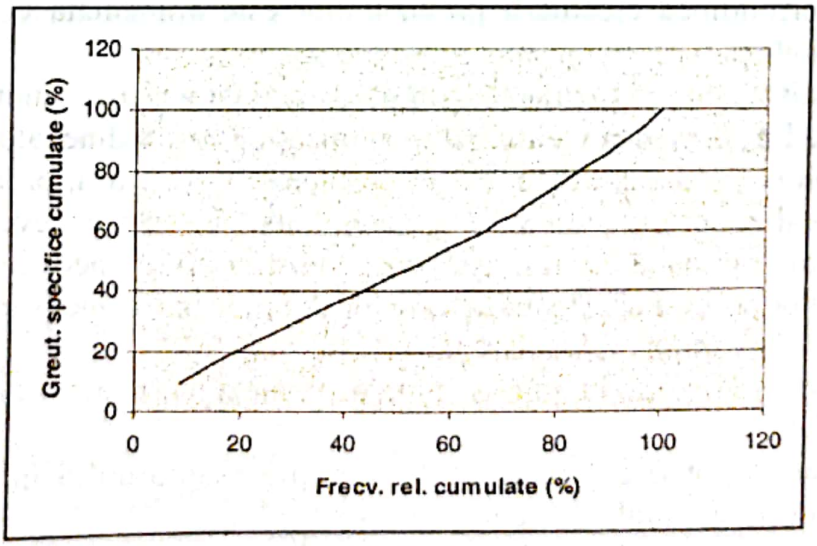


Figura 3.12. Curba Lorentz

Scanned with OKEN Scanner

CAPITOLUL 4
UTILIZAREA SONDAJULUI STATISTIC ÎN STUDIUL FENOMENELOR ECONOMICO-SOCIALE

4.1. Introducere. Definiție, avantaje și dezavantaje ale sondajului statistic

De multe ori, populațiile statistice sunt atât de numeroase, încât realizarea unei observări statistice totale – ce presupune înregistrarea variabilelor de la toate unitățile colectivității – este prea costisitoare de resurse materiale, financiare, umane, de timp etc. În astfel de cazuri se recurge la o *metodă de observare parțială – la sondajul (sau selecția) statistică, prin selectarea unei părți a populației totale.* În urma înregistrării caracteristicilor statistice de la unitățile selectate și a prelucrării datelor obținute, este necesară extinderea rezultatelor la nivelul întregii colectivități, astfel încât acestea să ne ofere o imagine pertinentă, validă, reprezentativă, asupra situației existente în colectivitatea generală.

Datorită complexității fenomenelor economico-sociale și a condițiilor de incertitudine care însoțesc desfășurarea acestora, se impune utilizarea unor metode de măsurare a probabilității de producere a unor evenimente. Chiar dacă dispunem de date complete pentru descrierea unui fenomen sau proces, există foarte puține evenimente predictibile în condiții de certitudine, cu probabilitate de 100%. De multe ori, incertitudinea efectuării previziunilor este alimentată chiar de incompletitudinea datelor.

Dacă dispunem de o colecție completă de date despre o anumită situație și putem, pe baza lor, prevedea un anumit eveniment cu certitudine, atunci acesta este un *mod deterministic* de abordare a unei probleme. Dacă, însă, producerea evenimentului se realizează cu o anumită probabilitate (nu este un eveniment sigur), atunci problema se pune în *termeni probabilistici*, fiind necesară identificarea tuturor scenariilor posibile și calcularea șanselor de producere a unei variante.

Avantajele utilizării sondajului statistic sunt:

- asigură o importantă economie de resurse materiale, financiare, umane și de timp;
- datele pot fi colectate de la unitățile eșantionului în mod operativ, într-un interval mai scurt de timp;
- obținerea rezultatelor se poate produce, de asemenea, într-un interval mai redus de timp;

Scanned with OKEN Scanner

- prin sondaj se asigură un volum suficient de informaţii, cu un grad de exactitate acceptabil;
- programul cercetării prin sondaj poate cuprinde mai multe variabile, care să asigure o caracterizare mai detaliată, sub multiple faţete, a populaţiei totale şi parţiale;
- poate fi folosit în domenii în care utilizarea observării totale este exclusă, întrucât în urma acestui proces se distrug unităţile statistice (de exemplu, în controlul calităţii unor produse).

Principalul **dezavantaj** al sondajului este legat de faptul că estimarea obţinută pe baza datelor din eşantion nu reflectă cu certitudine situaţia reală din colectivitatea generală, ci doar cu o anumită probabilitate (care dă siguranţa estimării). Însă dacă formarea eşantionului se produce în condiţiile respectării principiilor selecţiei aleatorii, se poate garanta, cu un anumit nivel de încredere, reprezentativitatea eşantionului şi erorile obţinute în urma aplicării sondajului statistic vor fi mai reduse.

4.2. Noţiuni specifice sondajului statistic

Submulţimea de unităţi extrasă dintr-o populaţie totală, de la care se înregistrează datele necesare efectuării investigaţiei statistice se numeşte ***eşantion***, ***colectivitate de selecţie*** sau ***colectivitate parţială***.

Se notează cu N *volumul colectivităţii totale* şi cu n *volumul eşantionului*, între N şi n existând relaţia:

$$1 \le n < N \qquad (4.1.)$$

Dacă colectivitatea parţială şi cea totală au fost structurate în *r* grupe după variaţia unei variabile statistice, atunci vom avea:

- Pentru colectivitatea totală: $N = \sum_{i=1}^{r} N_i$, unde N_i este volumul grupei *i*;
- Pentru eşantion: $n = \sum_{i=1}^{r} n_i$, unde n_i este volumul grupei *i*;

Principalul motiv pentru care se recurge atât de des la studierea unui eşantion şi nu al unei populaţii totale este costul (în resurse financiare şi de timp). Cu ajutorul inferenţei statistice putem estima un parametru al populaţiei generale bazându-ne pe datele unui eşantion, de volum mult mai mic decât al colectivităţii totale. De exemplu, dacă managerul unui post TV doreşte să afle ponderea telespectatorilor care urmăresc programele postului său, deoarece numărul total al telespectatorilor este foarte mare, determinarea proporţiei dorite este, practic,

Scanned with OKEN Scanner

imposibilă şi foarte costisitoare printr-o cercetare totală. De aceea se poate selecta un eşantion de telespectatori pe baza căruia se poate calcula ponderea dorită.

Indicatorul statistic obţinut pe baza prelucrării datelor eşantionului poate fi foarte aproape de parametrul pe care dorim să-l estimăm, al populaţiei totale, dacă populaţia-ţintă (populaţia despre care vrem să facem inferenţa) şi populaţia supusă eşantionării (populaţia din care selectăm eşantionul) sunt aceleaşi. În practică, între ele pot exista diferenţe. De exemplu, dacă vrem să determinăm proporţia telespectatorilor din toată ţara care urmăresc programele postului TV respectiv şi selectăm un eşantion dintre telespectatorii din Bucureşti, atunci populaţia ţintă şi populaţia eşantionată nu va fi aceeaşi, iar rezultatele ce se vor obţine pe baza sondajului nu vor reflecta situaţia din întreaga ţară. Pentru ca acest lucru să nu se întâmple, trebuie ca eşantionul să includă telespectatorii bărbaţi şi femei, de diferite vârste, venituri, ocupaţii şi zone de reşedinţă, în proporţii asemănătoare cu cele din populaţia ţintă. Acest lucru înseamnă că eşantionul selectat trebuie să fie *reprezentativ*, adică să reunească aceleaşi trăsături esenţiale ca şi populaţia generală din care este extras. În acest fel, eroarea obţinută în cazul sondajului va fi ţinută sub control şi diminuată.

Probabilitatea de producere a unui eveniment reprezintă frecvenţa relativă cu care este aşteptată apariţia evenimentului. Probabilitatea de producere a unui eveniment poate fi obţinută în 3 moduri:

a) empiric;
b) teoretic;
c) subiectiv.

Probabilitatea empirică sau **experimentală** reprezintă frecvenţa relativă observată a apariţiei evenimentului pe termen lung.

De exemplu, dacă am aruncat cu zarul de 30 de ori şi de 4 ori a ieşit numărul şase, atunci probabilitatea de apariţie a numărului şase a fost de $\frac{4}{30} \cdot \frac{2}{15} = 0{,}13$.

Dacă un experiment se poate produce în n moduri, fiecare cu şanse egale de apariţie, probabilitatea de producere a unui dintre ele este de 1/n (**probabilitate teoretică**). De exemplu, un zar are 6 feţe. Probabilitatea apariţiei numărului 3 este de 1/6.

În unele situaţii, rezultatele unui experiment nu au şanse egale de producere; de exemplu, vrem să obţinem un anumit profit în afacerea pe care am demarat-o la sfârşitul anului în curs. Probabilitatea de producere a acelui eveniment este determinată printr-o apreciere **subiectivă**, ca reflectând gradul în care credem noi că va apare varianta dorită (o evaluare personală, subiectivă a situaţiei, bazată deseori pe intuiţie).

Un indicator statistic determinat pe baza datelor referitoare la o colectivitate totală şi care caracterizează repartiţia statistică se numeşte ***parametru***. De

Scanned with OKEN Scanner

cele mai multe ori, însă, parametrul este un număr fix, constant, dar necunoscut, deoarece nu se cunosc datele pentru toate unităţile colectivităţii generale.

Indicatorul statistic determinat pe baza datelor culese de la unităţile eşantionului se numeşte ***estimator*** (sau ***statistică***). Cum un eşantion de volum *n* poate fi format aleatoriu în mai multe moduri dintr-o colectivitate generală de volum N, atunci din fiecare eşantion astfel format se poate calcula câte o statistică. Aşadar, statistica (estimatorul) nu are un nivel fix, constant, ci variază ca valoare de la un eşantion la altul, fiind de fapt o variabilă aleatorie.

Diferenţa dintre nivelul estimatorului calculat pe baza datelor din eşantion şi nivelul parametrului din colectivitatea generală se numeşte ***eroare de estimaţie***. Deoarece, de regulă, valoarea parametrului este necunoscută, nici eroarea de estimaţie nu se poate determina.

Pentru principalii indicatori statistici determinaţi pe baza datelor de la nivelul populaţiei totale sau al eşantionului, vom adopta notaţii diferite:

- **Media:**

  a) *pentru o caracteristică nealternativă*

  - la nivelul populaţiei totale, vom nota media cu μ şi se va determina cu relaţia:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{(pentru o serie simplă)} \qquad (4.2.)$$

Sau

$$\mu = \frac{\sum_{i=1}^{r} x_i N_i}{\sum_{i=1}^{r} N_i} \quad \text{(pentru o distribuţie de frecvenţe)} \qquad (4.3.)$$

  - la nivelul eşantionului vom nota media cu $\bar{x}$, dedusă cu relaţia:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{(pentru o serie simplă)} \qquad (4.4.)$$

Sau

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} \quad \text{(pentru o distribuţie de frecvenţe)} \qquad (4.5.)$$

Scanned with OKEN Scanner

b) *pentru o caracteristică alternativă*

- media caracteristicii alternative la nivelul populației generale (π) se definește ca raport între numărul unităților care posedă caracteristica din întreaga colectivitate (M) și numărul total de unități din colectivitatea generală (N), adică:

$$\pi = \frac{M}{N} \qquad (4.6.)$$

- media caracteristicii alternative la nivelul eșantionului (p) se definește ca raport între numărul unităților din eșantion care posedă caracteristica (m) și volumul eșantionului (n), adică:

$$p = \frac{m}{n} \qquad (4.7.)$$

- **Dispersia**

c) *pentru o caracteristică nealternativă*

- la nivelul colectivității generale, notăm dispersia caracteristicii analizate cu σ^2, determinată cu relația:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \text{ (pentru o serie simplă)} \qquad (4.8.)$$

Sau

$$\sigma^2 = \frac{\sum_{i=1}^{r}(x_i - \mu)^2 N_i}{\sum_{i=1}^{r} N_i} \text{ (pentru o serie de frecvențe)} \qquad (4.9.)$$

- la nivelul eșantionului, dispersia se va nota cu s^2 și se va determina conform relației:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \text{ (pentru o serie simplă)} \qquad (4.10.)$$

Sau

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i} \text{ (pentru o serie de frecvențe)} \qquad (4.11.)$$

Scanned with OKEN Scanner

Formula (4.10.) sau (4.11.) a dispersiei se foloseşte dacă eşantionul este de volum mare (n > 30). În cazul unui eşantion de volum redus (n < 30), dispersia se va calcula cu relaţia:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \text{ (pentru o serie simplă)} \qquad (4.12.)$$

Sau

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i - 1} \text{ (pentru o serie de frecvenţe)} \qquad (4.13.)$$

Calcularea dispersiei la nivelul eşantionului de volum redus, prin relaţia (4.12.) sau (4.13.) permite obţinerea unei estimaţii mai bune a dispersiei caracteristicii în colectivitatea generală (spunem că dispersia calculată pe baza relaţiei (4.12.) sau (4.13.) este un estimator nedeplasat al dispersiei din colectivitatea generală). Aceasta înseamnă că media dispersiilor determinate pentru toate eşantioanele de volum n posibil a fi formate va fi egală cu dispersia variabilei în colectivitatea generală.

d) *pentru o caracteristică alternativă*

- la nivelul colectivităţii generale, notăm dispersia cu σ_w^2 şi se va determina conform relaţiei:

$$\sigma_w^2 = \pi(1-\pi) \qquad (4.14.)$$

- la nivelul eşantionului, dispersia se va nota cu s_w^2 şi se va determina conform relaţiei:

$$s_w^2 = p(1-p) \qquad (4.15.)$$

4.3. Modalităţi de extracţie aleatorie a eşantioanelor. Tipuri de sondaj

Există trei tipuri de sondaj:

a) sondaj aleatoriu (probabilist);
b) sondaj dirijat;
c) sondaj mixt.

Scanned with OKEN Scanner

Formarea în mod aleatoriu a eşantioanelor este metoda utilizată cel mai des în efectuarea unui sondaj, întrucât, dacă se respectă principiile selecţiei aleatorii, se poate asigura reprezentativitatea eşantionului şi se pot asigura erori de selecţie mai mici. Selecţia dirijată ia în considerare anumite elemente preferenţiale în formarea eşantioanelor, în timp ce selecţia mixtă combină principiile selecţiei aleatorii cu cele ale selecţiei dirijate.

În cadrul sondajului aleatoriu există mai multe subtipuri:

- sondaj simplu;
- sondaj tipic (stratificat);
- sondaj de serii (în cuiburi);
- sondaj în trepte;
- sondaj secvenţial.

A) ***Eşantionarea aleatorie simplă*** este cea în care toate unităţile ce formează populaţia totală au şanse egale şi nenule de a fi incluse în eşantion. Principalele modalităţi în care putem realiza acest lucru sunt:

1. **Procedeul loteriei (schema urnei cu bile)** ce constă în următorii paşi:

P1 – fiecărei unităţi a populaţiei totale i se alocă un cod (număr), care identifică în mod unic elementul respectiv. Pot fi folosite drept coduri seria şi numărul cărţii de identitate, numărul asigurării medicale, numărul legitimaţiei de serviciu, numărul carnetului de şofer etc.;

P2 – se trec codurile respective pe bile/cartonaşe/bileţele, amestecându-se într-o urnă. În total, în urnă sunt, în momentul iniţial, N bileţele (N = volumul populaţiei totale);

P3 – se extrage un cartonaş din urnă, la întâmplare şi se citeşte codul înscris pe el; unitatea cu codul respectiv va fi inclusă în eşantion. Pot exista în continuare două situaţii:

I. cartonaşul sau bila extrasă este pusă înapoi în urnă, participând la o nouă extragere (selecţie repetată, ce corespunde schemei bilei revenite); aceasta înseamnă că o unitate poate fi selectată, prin jocul întâmplării, de mai multe ori în eşantion. Probabilitatea ca o bilă să fie inclusă în eşantion, la fiecare extragere, este de $1/N$. Se repetă procedeul până când sunt incluse n unităţi în eşantion (n = volumul eşantionului). Prin această modalitate pot fi formate N^n eşantioane diferite, de volum n, dintr-o populaţie totală de N unităţi.

II. Cartonaşul sau bila extrasă nu mai este reintrodusă în urnă (selecţie nerepetată, ce corespunde schemei bilei nerevenite); în acest fel, o unitate poate fi inclusă în eşantion o singură dată, acest gen de sondaj generând erori mai mici decât sondajul repetat. La a doua extragere vor participa $(N - 1)$ bile rămase, la a treia extragere $(N - 2)$ bile rămase etc., la extragerea i vor participa $N - (i - 1)$ bile rămase în urnă, procedeul repetându-se până când se completează cele n unităţi în eşantion.

Scanned with OKEN Scanner

Probabilitatea ca o unitate să fie inclusă în eşantion, la extragerea i este de $\frac{1}{N-(i-1)}$, prin acest procedeu putându-se forma C_N^n eşantioane diferite de volum n dintr-o populaţie totală de N unităţi.

2. Procedeul tabelului cu numere aleatorii

Tabelul cu numere aleatorii este o listă de numere în care cifrele de la 0 la 9 apar într-o ordine aleatorie, fiecare cu o probabilitate de 1/10, în mod independent unele de altele. Utilizarea tabelului cu numere aleatorii în formarea eşantioanelor presupune parcurgerea următorilor paşi:

P1 – se alocă fiecărei unităţi a populaţiei totale un cod numeric (număr de ordine) ce identifică în mod unic unitatea respectivă; codurile vor fi de la 1 la N;

P2 – se stabileşte aleatoriu poziţia din tabel începând de la care se efectuează citirea numerelor (linia şi coloana);

P3 – de la poziţia stabilită se citesc şiruri de cifre de la stânga la dreapta şi de sus în jos, într-o direcţie firească;

P4 – şirul de cifre citit se împarte în secvenţe de câte k cifre, unde k este numărul de cifre conţinut în N; aceste secvenţe vor reprezenta numerele de ordine ale unităţilor ce vor fi incluse în eşantion;

P5 – dintre secvenţele de câte k cifre astfel formate, se elimină numerele de 0 şi cele mai mari decât N, deoarece ele nu corespund nici unei unităţi din colectivitatea totală;

P6 – secvenţele rămase în urma eliminării indică numerele de ordine ale unităţilor din eşantion.

P7. Se continuă citirea până când se găsesc n secvenţe numerice valide pentru formarea completă a eşantionului. Dacă selecţia este repetată, în cazul în care o secvenţă validă de cifre apare de mai multe ori, atunci unitatea respectivă va fi inclusă în eşantion de acelaşi număr de ori. Dacă selecţia este nerepetată, atunci unitatea respectivă va fi inclusă o singură dată în eşantion.

Exemplul 4.1. Din cei N = 90 de salariaţi ai unei mari companii se doreşte formarea unui eşantion de n = 10 salariaţi. Se stabileşte ca citirea cifrelor din tabelul cu numere aleatorii să înceapă de la rândul 5, coloana 8. Se vor citi numerele: 6295, 5658, 1091, 5485, 7888, 5791, 1169.

K = 2

Rezultă următoarele grupuri de câte 2 cifre:

62, 95, 56, 58, 10, 91, 54, 85, 78, 88, 57, 91, 11, 69. Se elimină numerele: 95, 91, 91. Rămân valide următoarele numere: 62, 56, 58, 10, 54, 85, 78, 88, 57, 11. Salariaţii cu codurile numerice respective vor fi incluşi în eşantion.

Scanned with OKEN Scanner

3. **Procedeul mecanic**

Formarea eşantionului prin procedeul mecanic presupune parcurgerea următorilor paşi:

P1 – se alocă unităţilor populaţiei totale numere de ordine după un criteriu independent de scopul cercetării statistice prin sondaj; fiecare număr de ordine identifică în mod unic o unitate statistică din baza de sondaj (de la 1 la N);

P2 – se determină un pas de numărare, prin împărţirea volumului colectivităţii totale la cel al eşantionului:

$$k = \frac{N}{n} \qquad (4.16.)$$

P3 – se stabileşte aleatoriu poziţia (numărul de ordine) de la care se începe extragerea unităţilor în eşantion: un număr între 1 şi k;

P4 – începând de la poziţia aleasă, se extrag succesiv unităţile cu numere de ordine aflate la distanţa k unele de altele (de exemplu, dacă N = 200, n = 10, k = 200/10 = 20; dacă începem citirea de la poziţia 6, se vor include în eşantion unităţile cu numerele de ordine: 6, 6 + 20 = 26, 26 + 20 = 46, 46 + 20 = 66 etc. până când se completează cele 10 unităţi din eşantion).

B) ***Eşantionarea aleatorie stratificată***

Pentru a obţine în urma unei cercetări selective rezultate cât mai exacte, este necesar să obţinem cât mai multe informaţii despre eşantion. Pentru aceasta se poate folosi eşantionarea aleatorie stratificată. Eşantionarea aleatorie stratificată presupune împărţirea populaţiei totale în straturi separate şi extragerea din fiecare strat a câte unui eşantion aleatoriu simplu. Salariaţii unei firme pot fi împărţiţi, de exemplu, după starea civilă (căsătorit/necăsătorit), sex (masculin/feminin), salariul (sub medie/peste medie), ocupaţie (economist/jurist/inginer etc.), statutul de angajat permanent sau colaborator, vechimea în muncă (sub 10 ani/10-15 ani/15-20 ani/peste 20 de ani). La fel, locuitorii unei regiuni pot fi şi ei stratificaţi după sex (masculin/feminin), mediul de proveniență (urban/rural), vârstă (0-10 ani/10-20 ani/20-30 ani/30-40 ani/40-50 ani/peste 50 ani) etc. Straturile nu trebuie să conţină suprapuneri, astfel că o unitate trebuie să fie inclusă, conform criteriului de stratificare, într-un singur strat/grupă. Calitatea inferenţei statistice poate fi optimizată dacă se reduce variaţia la nivelul eşantionului. Astfel, dacă formăm straturi omogene, din unităţi asemănătoare din punctul de vedere al caracteristicii urmărite, atunci variaţia din cadrul fiecărui strat va fi mai mică şi, prin efectul regulii de adunare a dispersiilor şi variaţia totală a caracteristicii la nivelul eşantionului se va diminua. De asemenea, este posibilă efectuarea unei inferenţe statistice şi pentru fiecare strat în parte.

Scanned with OKEN Scanner

C) ***Eşantionarea tip cluster (în cuiburi)*** presupune formarea unor eşantioane aleatorii simple de grupuri sau clustere de unităţi.

Acest tip de eşantionare se poate folosi atunci când determinarea unei liste complete a unităţilor populaţiei generale este dificil de realizat şi implică costuri foarte ridicate (spre exemplu, atunci când unităţile populaţiei totale sunt dispersate teritorial). De exemplu, dorim să estimăm cheltuielile medii lunare pe produse alimentare ale unei gospodării dintr-un oraş mare. Dacă aplicăm o eşantionare simplă, vom avea nevoie de o listă completă a tuturor gospodăriilor din acel oraş, operaţiune costisitoare deoarece acestea au un grad foarte mare de răspândire în spaţiu. Dacă vrem să formăm eşantionul în mod stratificat, avem nevoie atât de o listă completă a tuturor gospodăriilor din oraş, cât şi de înregistrarea unei alte variabile în funcţie de care se vor grupa gospodăriile pe straturi (de exemplu, gruparea gospodăriilor după venitul mediu al acestora). Pentru reducerea costurilor implicate de formarea eşantionului, se poate considera un cluster format dintr-un cartier şi care va reprezenta toate gospodăriile din acel cartier. Se va forma un eşantion de cartiere, selectat aleatoriu şi poate fi înregistrat nivelul cheltuielilor cu produse alimentare pentru fiecare gospodărie din cadrul fiecărui cluster (reducerea distanţelor parcurse de personalul de specialitate duce la limitarea costurilor).

4.4. Posibilităţi de estimare a parametrilor colectivităţii generale

Estimarea parametrilor populaţiei totale pe baza inferenţei statistice se poate face în două moduri:

- prin *estimare punctuală*, în care parametrul corespunzător colectivităţii totale este estimat prin indicatorul statistic calculat la nivelul eşantionului (Figura 4.1.);
- prin *estimare pe un interval*, în care parametrul aferent colectivităţii totale este estimat pe baza unui interval de încredere, care reflectă cât de mult variază mediile diverselor eşantioane de volum *n* formate dintr-o populaţie totală de volum N (Figura 4.2.).

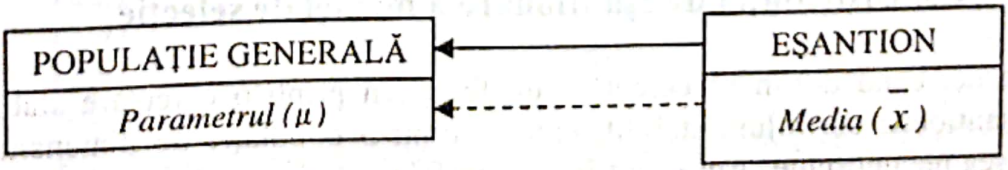


Figura 4.1. Estimare punctuală a parametrului populaţiei generale

Scanned with OKEN Scanner

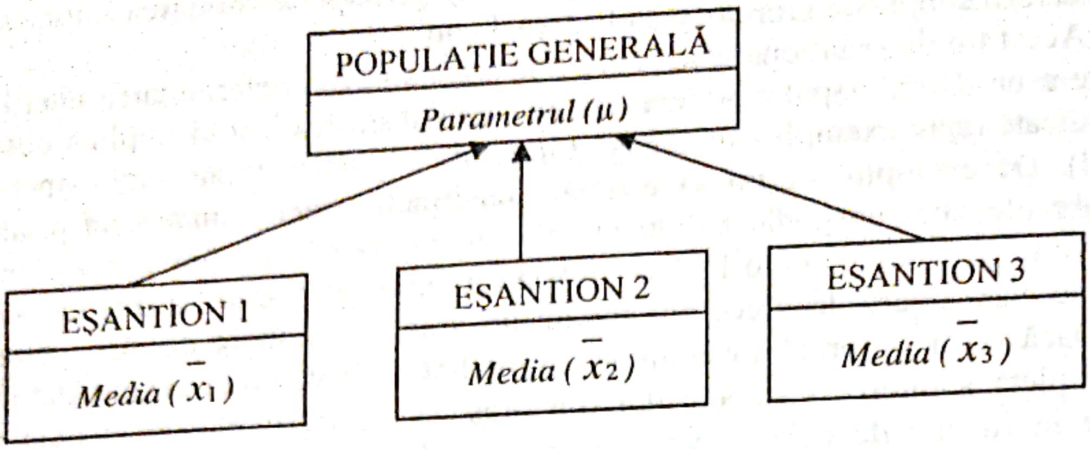


Figura 4.2. Estimare pe bază de interval de încredere a parametrului populației generale

În cazul estimării punctuale, se pune întrebarea dacă indicatorul statistic determinat pe baza datelor dintr-un eșantion este un bun estimator al parametrului populației generale. În cazul în care nu cunoaștem valoarea parametrului din colectivitatea generală, pentru determinarea acestuia, se impune utilizarea unor indicatori statistici aferenți eșantionului. Fiecăruia dintre indicatorii statistici prezentați în capitolele anterioare, precum și în cele ce vor urma, le corespunde un parametru din colectivitatea generală, iar fiecare din acești indicatori statistici ai eșantionului reprezintă un estimator al parametrului din colectivitatea generală. Acești estimatori nu estimează în mod exact nivelul parametrului din populația totală, ci pot include unele erori. Mărimea erorii efectuate prin estimarea punctuală nu este posibil de determinat, deoarece de cele mai multe ori nu se cunoaște valoarea parametrului.

O formă de inferență statistică o constituie testarea ipotezelor, care presupune compararea rezultatelor sondajului cu valorile cunoscute sau presupuse ale parametrilor colectivității generale.

4.5. Distribuția de eșantionare a mediei de selecție

Atunci când dorim să cercetăm un fenomen printr-o cercetare statistică, folosim eșantioane de volum stabilit, extrase dintr-o populație de dimensiune și formă adesea necunoscute. Am văzut în paragrafele anterioare că, pentru a efectua o inferență asupra populației totale trebuie să utilizăm un indicator al statisticii descriptive care să caracterizeze anumite variabile pentru un eșantion. Se pune problema siguranței acestor estimări și în ce măsură rezultatele obținute pe baza eșantionului pot fi generalizate la nivelul întregii colectivități. Dacă vom continua

Scanned with OKEN Scanner

să formăm şi alte eşantioane de acelaşi volum, unii dintre indicatorii obţinuţi vor fi mai mari decât parametrul populaţiei totale, iar alţii – mai mici.

Vom considera că parametrul ce trebuie estimat este ***media***. Distribuţia de eşantionare este tot o distribuţie de probabilitate teoretică a unor valori posibile ale unor indicatori statistici, ce ar putea apărea dacă am forma un număr infinit de eşantioane de volum dat, dintr-o populaţie totală. Între distribuţiile de frecvenţe şi distribuţiile de eşantionare există unele diferenţe:

- În ceea ce priveşte modul de obţinere al fiecărui tip de distribuţie: în timp ce o distribuţie de frecvenţe este obţinută în mod empiric, prin înregistrarea observaţiilor reale şi a frecvenţelor lor de apariţie, o distribuţie de eşantionare reprezintă o distribuţie teoretică. Aceasta deoarece în practică este imposibil să se formeze toate eşantioanele posibile de volum dat, pentru a determina, pentru fiecare eşantion, un estimator (medie) al parametrului colectivităţii generale; de aceea, pentru a descrie şi caracteriza forma unei distribuţii de eşantionare se folosesc teoreme matematice.
- În ceea ce priveşte tipul datelor analizate: în cazul distribuţiilor de frecvenţe sunt analizate valorile individuale ale unei variabile, pentru un estimator, în timp ce într-o distribuţie de eşantionare analizăm o serie de statistici (indicatori statistici) obţinuţi pentru diferite eşantioane. Putem avea distribuţii de eşantionare pentru orice indicator statistic: medie, abatere medie pătratică, proporţii, corelaţii şi alţii, dar în continuare vom studia distribuţiile de eşantionare ale mediei de selecţie şi le vom descrie matematic.

Să considerăm că de la cele N = 4 unităţi ale unei populaţii totale, vrem să formăm eşantioane de volum $n = 2$ unităţi. Valorile unei caracteristici X, înregistrate de la cele 4 unităţi ale populaţiei generale sunt: $x_1 = 2$; $x_1 = 3$; $x_1 = 4$; $x_1 = 5$. Dacă aplicăm o selecţie repetată, atunci se pot forma cel mult 16 eşantioane de câte două unităţi statistice:

(2;2)	(3;2)	(4;2)	(5;2)
(2;3)	(3;3)	(4;3)	(5;3)
(2;4)	(3;4)	(4;4)	(5;4)
(2;5)	(3;5)	(4;5)	(5;5)

Probabilitatea de apariţie a unui eşantion va fi de 1/16 = 0,0625. Pentru fiecare eşantion se determină câte o medie (Tabelul 4.1.)

Scanned with OKEN Scanner

Tabelul 4.1.

Eşantionul (i)	Media eşantionului $\bar{x}_i$
(2;2)	2
(2;3)	2,5
(2;4)	3
(2;5)	3,5
(3;2)	2,5
(3;3)	3
(3;4)	3,5
(3;5)	4

Eşantionul (i)	Media eşantionului $\bar{x}_i$
(4;2)	3
(4;3)	3,5
(4;4)	4
(4;5)	4,5
(5;2)	3,5
(5;3)	4
(5;4)	4,5
(5;5)	5

Distribuţia de frecvenţe a mediilor de selecţie este prezentată în Tabelul 4.2., iar poligonul frecvenţelor alcătuit pe baza datelor din Tabelul 4.2. se regăseşte în Figura 4.3.

Tabelul 4.2.

Media de selecţie	Frecvenţe absolute (n_i)	Frecvenţe relative (n_i^*)
2,0	1	0,0625
2,5	2	0,125
3,0	3	0,1875
3,5	4	0,25
4,0	3	0,1875
4,5	2	0,125
5,0	1	0,0625
Total	16	1,00

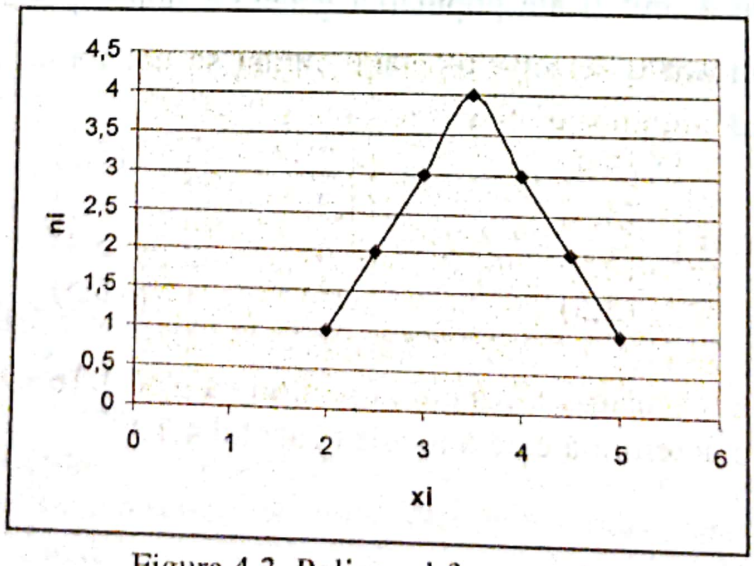


Figura 4.3. Poligonul frecvenţelor

Scanned with OKEN Scanner

Distribuţia mediilor de selecţie seamănă cu o distribuţie normală, aproape simetrică faţă de medie.

Din păcate, în practică nu se procedează la alcătuirea tuturor eşantioanelor de volum fix posibile, ci se lucrează doar cu un singur eşantion, fiind disponibile doar datele colectate de la unităţile acestui eşantion. Întrucât nu se cunosc, deci, toate mediile eşantioanelor posibil a fi alcătuite, nu se poate obţine distribuţia de eşantionare a mediilor de selecţie. Totuşi, această distribuţie poate fi caracterizată cu ajutorul a doi indicatori: media şi abaterea medie pătratică a mediilor de selecţie. De fapt, o distribuţie de valori este definită sub aspect matematic prin perechea de indicatori formată din media şi abaterea medie pătratică.

a) Media distribuţiei de eşantionare a mediilor de selecţie

Se observă din exemplul numeric anterior că mediile tuturor eşantioanelor posibil a fi formate sunt normal distribuite. Într-adevăr, conform teoremei limită centrală distribuţia normală descrie cel mai bine o distribuţie de eşantionare a mediei de selecţie. Mai exact, teorema limită centrală stipulează următoarele: dacă mai multe eşantioane de volum n sunt extrase aleatoriu dintr-o populaţie de volum N, indiferent de forma distribuţiei colectivităţii totale, pe măsură ce n este tot mai mare, distribuţia mediilor de selecţie tinde spre o distribuţie normală; totodată şi media mediilor de selecţie tinde să se apropie de media caracteristicii la nivelul colectivităţii totale, adică:

$$\mu(\bar{x}_i) = \mu \qquad (4.17.)$$

Într-adevăr, în exemplul numeric anterior, s-a obţinut media la nivelul colectivităţii totale:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{2+3+4+5}{4} = 3{,}5$$

iar media mediilor de selecţie:

$$\mu(\bar{x}_i) = \frac{\sum_{i=1}^{r} x_i n_i}{\sum_{i=1}^{r} n_i} = \frac{2\cdot 1 + 2{,}5\cdot 2 + 3\cdot 3 + 4\cdot 3{,}5 + 4\cdot 3 + 4{,}5\cdot 2 + 5\cdot 1}{16} = 3{,}5 = \mu$$

Spunem, aşadar, că media de selecţie este un estimator nedeplasat al mediei colectivităţii totale, dacă media mediilor de selecţie este egală cu media colectivităţii generale (adică cu parametrul ce trebuie estimat).

Pentru ca teorema limită centrală să fie adevărată, se fac următoarele presupuneri:

- media şi abaterea medie pătratică a populaţiei sunt fixe;

Scanned with OKEN Scanner

- eşantioanele sunt formate în mod aleatoriu, fiecare unitate a populaţiei totale având şanse egale de a fi inclusă în eşantion;
- toate eşantioanele au acelaşi număr de observaţii.

Din teorema limită centrală rezultă că indiferent de forma distribuţiei pentru populaţia ce va fi eşantionată (normal distribuită, asimetrică, bimodală, în formă de U, L sau de I), distribuţia de eşantionare va fi întotdeauna normală şi cu cât volumul eşantionului va fi mai mare ($n \geq 30$), cu atât distribuţia de eşantionare se apropie de cea normală.

Principiile enunţate în teorema limită centrală stau la baza determinării erorii medii de reprezentativitate şi a aplicării inferenţei statistice de la nivelul eşantionului la nivelul colectivităţii generale.

b) *Abaterea medie pătratică a distribuţiei de eşantionare a mediilor de selecţie* este egală cu abaterea medie pătratică din populaţia totală, micşorată de $\sqrt{n}$ ori:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad (4.18.)$$

Indicatorul arată cum variază mediile de selecţie ale diferitelor eşantioane faţă de media colectivităţii generale şi poartă numele de ***eroare medie de reprezentativitate***. Altfel spus, dispersia mediilor de selecţie în jurul mediei colectivităţii generale este de n ori mai mică decât dispersia la nivelul colectivităţii generale.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad (4.19.)$$

Cum de cele mai multe ori dispersia colectivităţii generale (σ^2) şi abaterea medie pătratică din colectivitatea generală (σ) nu sunt cunoscute, ele se aproximează prin indicatorii pereche de la nivelul eşantionului: dispersia (s^2) şi abaterea medie pătratică (s). Aşadar, dispersia mediilor de selecţie în jurul parametrului (μ) va fi estimată prin:

$$s_{\bar{x}}^2 = \frac{s^2}{n} \qquad (4.20.)$$

şi abaterea medie pătratică a mediilor de selecţie în jurul mediei (μ) va fi estimată prin:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (4.21.)$$

Scanned with OKEN Scanner

Aşadar, teorema limită centrală descrie forma distribuţiei de eşantionare a mediilor de selecţie, arătând, totodată, legătura dintre media μ a colectivităţii generale şi media distribuţiei de eşantionare $\mu(\bar{x}_i)$, pe de o parte, şi relaţia dintre abaterea standard a populaţiei totale (σ) şi deviaţia standard a distribuţiei de eşantionare ($\sigma_{\bar{x}}$).

4.6. Sondajul simplu aleatoriu pentru o caracteristică nealternativă

4.6.1. Determinarea erorilor de sondaj şi a intervalului de încredere

Să presupunem că o mare companie de telefonie mobilă organizează un sondaj pe un segment de clienţi din Bucureşti privind un nou pachet de servicii, reieşind din respectiva cercetare că 35% dintre clienţi ar dori să încerce noua gamă de servicii. În aceste condiţii, ce s-ar putea spune despre toţi potenţialii clienţi existenţi în Bucureşti şi care ar fi atitudinea lor vizavi de noul pachet de servicii? Dar despre clienţii din toată ţara sau de pretutindeni? La o asemenea întrebare, s-ar putea obţine un răspuns sigur (cu probabilitate 100%) numai dacă s-ar organiza o observare statistică totală şi dacă toţi subiecţii intervievaţi ar răspunde în mod sincer. Rezultatele ar fi, însă, valabile numai la un anumit moment de timp şi anume la momentul de referinţă (momentul critic) la care se referă răspunsurile date. O asemenea operaţiune ar fi, însă, costisitoare, mare consumatoare de resurse financiare şi de timp. De cele mai multe ori, rezultatele trebuie obţinute într-un timp foarte scurt, în mod operativ, de aceea apelăm cel mai adesea la o cercetare statistică prin sondaj, care ne permite să extindem rezultatele obţinute pe baza datelor dintr-un eşantion la nivelul populaţiei totale, apelând la procesul de inferenţă statistică. Am văzut până acum că, în anumite condiţii, indicatorii statistici aferenţi colectivităţii de selecţie urmează o lege de repartiţie normală. Acurateţea rezultatelor derivate dintr-o cercetare selectivă depinde de următorii factori:

- mărimea eşantionului;
- variaţia variabilei urmărite din cadrul populaţiei;
- gradul/nivelul de încredere cu care dorim să garantăm rezultatele obţinute.

Modul în care cei trei factori îşi exercită influenţa asupra rezultatelor sondajului este următorul:

- rezultatele vor fi mai exacte (adică mai apropiate de rezultatele care s-ar fi obţinut dintr-o observare totală) cu cât volumul eşantionului creşte;
- cu cât variabila studiată prezintă o variabilitate mai mare în populaţia iniţială, cu atât trebuie să ne asigurăm că eşantionul reflectă această dispersare a valorilor caracteristicii şi de aceea trebuie să selectăm mai

Scanned with OKEN Scanner

multe unități în eșantion. Prin contrast, dacă într-o populație totală variabila studiată, în legătură cu care trebuie efectuată cercetarea statistică, nu prezintă deloc variație, manifestându-se o singură variantă la toate unitățile colectivității, atunci ar fi de ajuns prelevarea în eșantion a unei singure unități statistice;

- extinderea rezultatelor de la nivelul eșantionului la nivelul colectivității totale nu se poate face în condiții de certitudine absolută, altfel spus, concluziile emise dintr-o cercetare prin sondaj nu pot fi garantate 100%, ci întotdeauna există un risc de a fi greșite, inexacte, și diferite de rezultatele dintr-o eventuală observare totală. Ca atare, se impune specificarea gradului/nivelului de probabilitate cu care vrem să garantăm rezultatele sondajului.

Așa cum vom vedea în continuare, cu cât este mai mare probabilitatea cu care sunt garantate rezultatele sondajului, cu atât rezultatul va fi mai puțin exact.

Am văzut că dintr-o populație totală de volum N se pot extrage mai multe eșantioane diferite, de volum n, pentru fiecare eșantion putându-se determina câte o medie $\bar{x}_i$, $i=\overline{1,k}$ (k este numărul de eșantioane maxim posibil a fi extras: N^n pentru eșantionare repetată și C_N^n pentru eșantionare nerepetată). Cel mai probabil mediile de selecție, care reprezintă niște estimatori punctuali ai parametrului μ diferă unele de altele, ele mișcându-se într-un interval de forma: $\left(\bar{x}-er;\bar{x}+er\right)$, adică la cel mult o distanță egală cu eroarea de selecție, de o parte și de alta a mediei de selecție. Așadar, estimarea parametrului μ se face printr-un interval de încredere.

A. **Determinarea erorilor de sondaj, a intervalului de încredere pentru sondajul aleatoriu simplu repetat**

Am văzut că distribuția de eșantionare a mediilor de selecție este o distribuție normală, de medie egală cu μ și abaterea medie pătratică $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$ (***eroare medie de reprezentativitate***) (vezi Figura 4.4. – exemplu pentru o probabilitate de garantare a rezultatelor sondajului de 95%.

Scanned with OKEN Scanner

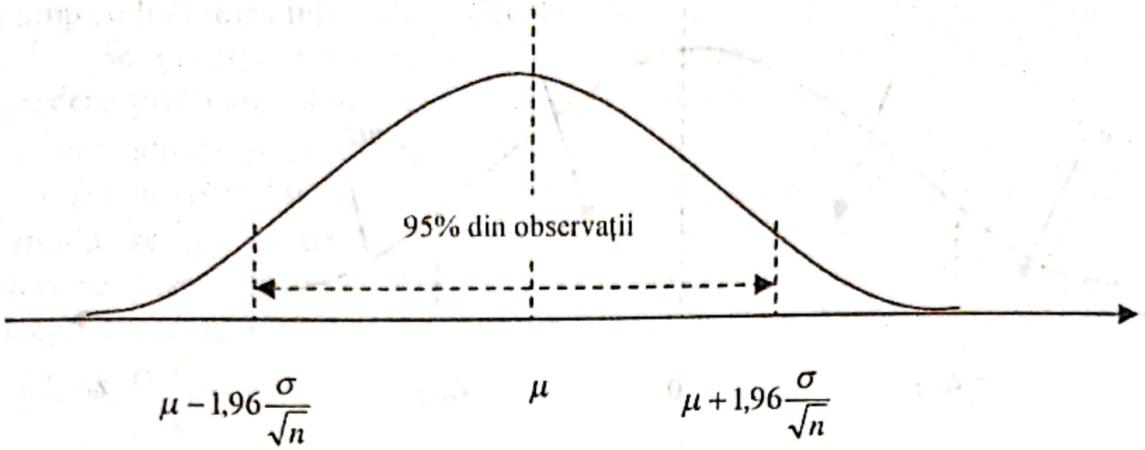


Figura 4.4. Distribuția de eşantionare a mediei de selecție

Pentru a găsi intervalul de încredere prin care estimăm parametrul μ al colectivității generale, pentru o variabilă nealternativă, se consideră variabila normal normată (redusă), de forma:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad (4.22.)$$

Am văzut că cel mai bun estimator al lui $\sigma_{\bar{x}}$ este $s_{\bar{x}}$ şi atunci variabila normal-normată va avea următoarea expresie:

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad (4.23.)$$

Variabila normal-normată z va avea o distribuție normală, de medie 0 şi dispersie 1 (Figura 4.5.).

Scanned with OKEN Scanner

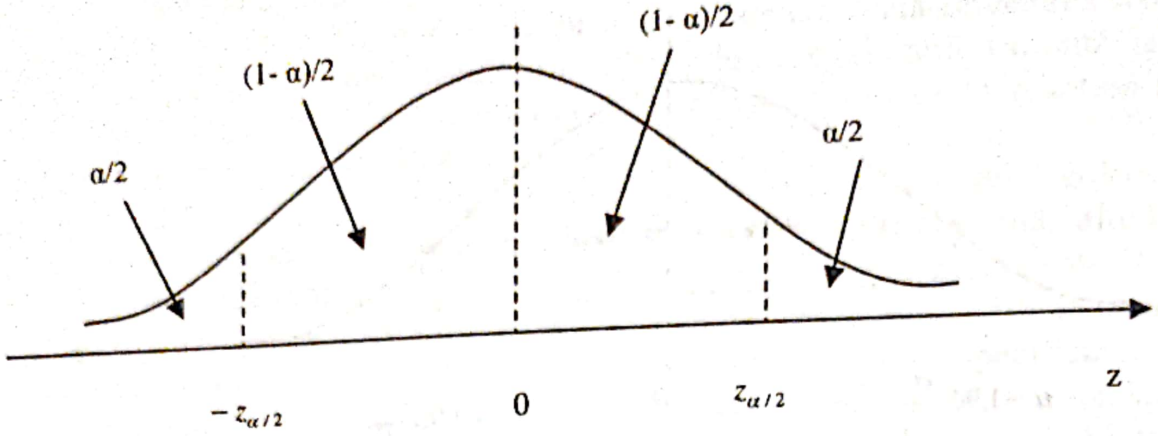


Figura 4.5. Distribuția variabilei normal-normate z

Intervalul de încredere pentru z este:

$$-z_{\alpha/2} \le z \le z_{\alpha/2} \tag{4.24.}$$

Probabilitatea ca variabila normal-normată z să se găsească în intervalul de mai sus este $100(1-\alpha)\%$, adică

$$P\left(-z_{\alpha/2} \le z \le z_{\alpha/2}\right) = 1-\alpha \tag{4.25.}$$

unde α este nivelul (pragul) de semnificație. Prin prelucrarea succesivă a expresiei (4.24.) se obține:

$$-z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2} \tag{4.26.}$$

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \tag{4.27.}$$

Produsul dintre variabila normal normată z şi eroarea medie de reprezentativitate se numeşte ***eroare limită maxim admisibilă*** ($\Delta_{\bar{x}}$):

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \tag{4.28.}$$

Relația (4.27.) devine, astfel:

$$\bar{x} - \Delta_{\bar{x}} \le \mu \le \bar{x} + \Delta_{\bar{x}} \tag{4.29.}$$

Probabilitatea cu care se garantează faptul că media la nivelul colectivității generale se găseşte în intervalul de încredere determinat este de $100(1-\alpha)\%$, unde α este nivelul (pragul) de semnificație. Înseamnă, deci, că rezultatul dat de relația (4.27.) nu este sigur 100%, existând o „neşansă" de $\alpha\%$ ca media μ a colectivității totale să nu se situeze în acest interval. Lungimea intervalului de încredere este

Scanned with OKEN Scanner

dublul erorii limită. Probabilitatea $100(1-\alpha)\%$ caracterizează *siguranța* estimării, în timp ce lungimea intervalului de încredere caracterizează *precizia* estimării.

Se observă că între volumul eşantionului şi lungimea intervalului de încredere prin care estimăm mărimea parametrului μ există o relaţie de inversă proporţionalitate şi astfel, cu cât creşte volumul eşantionului, cu atât se reduce lungimea intervalului de încredere şi estimarea va fi mai exactă (şi invers). De exemplu, se ştie că dacă o variabilă este normal distribuită, atunci 95% dintre valori se găsesc la o distanţă de 1,96 x abaterea medie pătratică a mediei de selecţie. Aşadar, 95% dintre valorile mediilor de selecţie se regăsesc în intervalul $\mu \pm 1{,}96 \cdot \sigma / \sqrt{n}$, adică

$$P\left(\mu - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0{,}95, \text{ adică} \qquad (4.30.)$$

$$\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \qquad (4.31.)$$

În practică, anumitor niveluri ale probabilităţii de garantare a rezultatelor sondajului le corespund anumite valori z (tabelate). Cele mai des utilizate niveluri de probabilitate sunt (Tabelul 4.3.):

Principalele niveluri ale probabilităţii utilizate în practica statistică

Tabelul 4.3.

Probabilitate	z
90%	1,645
95%	1,96
95,45%	2
99,73%	3

Cum precizia şi siguranţa estimării sunt invers proporţionale, în practică trebuie găsit un punct de echilibru între mărimea intervalului de încredere şi probabilitatea cu care parametrul μ al colectivităţii generale se regăseşte în intervalul de încredere estimat.

Pe lângă medie, se poate estima şi nivelul total al caracteristicii. Cum am văzut că media colectivităţii generale μ se defineşte conform relaţiei (4.2.), rezultă atunci:

$$\sum_{i=1}^{N} x_i = \mu \cdot N \qquad (4.32.)$$

şi deci intervalul de încredere pentru nivelul total al caracteristicii se obţine prin înmulţirea dublei inegalităţi din relaţia (4.29.) cu N:

Scanned with OKEN Scanner

$$N\left(\bar{x}-\Delta_{\bar{x}}\right)\leq\sum_{i=1}^{N}x_i\leq N\left(\bar{x}+\Delta_{\bar{x}}\right) \qquad (4.33.)$$

Exemplul 4.2. Pentru cei 10.000 de clienţi ai unei bănci se doreşte estimarea mărimii medii a creditelor contractate în valută în ultimul an. În urma selectării aleatorii simple repetate a 1000 de clienţi, a rezultat o valoare medie a creditului de 8270 de Euro, cu o abatere medie pătratică de 530 de Euro. Să se estimeze, cu o probabilitate de 95% ($z = 1{,}96$), valoarea medie a unui credit în Euro.

Rezolvare:

Se calculează eroarea medie de reprezentativitate:

$$\sigma_{\bar{x}}\approx s_{\bar{x}}=\frac{s}{\sqrt{n}}=\frac{530}{\sqrt{1000}}=16{,}76\text{ Euro}$$

Se determină eroarea limită:

$$\Delta_{\bar{x}}=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}=1{,}96\cdot 16{,}76=32{,}85\text{ Euro}$$

Se determină intervalul de încredere pentru valoarea medie a creditelor:

$$8270-32{,}85\leq\mu\leq 8270+32{,}85$$

$$8237{,}15\leq\mu\leq 8302{,}85\text{ Euro}$$

Dacă vrem să estimăm şi valoarea totală a creditelor, se înmulţeşte ultima dublă inegalitate cu 10.000, rezultând:

$$82.371.500\leq\sum_{i=1}^{N}x_i\leq 83.028.500\text{ Euro}$$

B. **Determinarea erorilor de sondaj, a intervalului de încredere pentru sondajul aleatoriu simplu nerepetat**

Sondajul aleatoriu simplu nerepetat diferă de cel repetat prin aceea că extragerile aleatorii ale unităţilor în eşantion sunt dependente unele de altele, întrucât o unitate poate fi inclusă în eşantion numai o dată, următoarea extragere efectuându-se din unităţile rămase în baza de sondaj şi care nu au fost selectate anterior. Această dependenţă între rezultatele mai multor extrageri succesive este exprimată prin *coeficientul de corecţie finită în populaţie* de forma: $\frac{N-n}{N-1}$. Pentru N foarte mare ($N > 500$), coeficientul de corecţie devine:

$$\frac{N-n}{N-1}\approx\frac{N-n}{N}=1-\frac{n}{N} \qquad (4.34.)$$

Scanned with OKEN Scanner

unde $\frac{n}{N}$ reprezintă fracţia de sondaj.

În cazul sondajului aleatoriu simplu nerepetat, eroarea medie de reprezentativitate va fi dată de:

$$\sigma_{\bar{x}} \approx s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{s^2}{n}\frac{N-n}{N-1}} \approx \sqrt{\frac{s^2}{n}\frac{N-n}{N-1}} \approx \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} \quad (4.35.)$$

Factorul $\frac{N-n}{N-1}$ este subunitar şi deci eroarea medie de reprezentativitate în cazul sondajului nerepetat este mai mică decât în cazul sondajului repetat. Pentru o fracţie de sondaj $\frac{n}{N} < 0{,}2$, factorul de corecţie se neglijează (tinde la 1).

Eroarea limită pentru sondajul simplu aleatoriu nerepetat este dată de:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sigma_{\bar{x}} = z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n}\frac{N-n}{N-1}} \approx z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} \quad (4.36.)$$

Intervalul de încredere pentru media colectivităţii generale va fi:

$$\bar{x} - \Delta_{\bar{x}} \le \mu \le \bar{x} + \Delta_{\bar{x}} \quad (4.37.)$$

adică

$$\bar{x} - z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} \le \mu \le \bar{x} + z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} \quad (4.38.)$$

Intervalul de încredere pentru nivelul total al caracteristicii:

$$N\left(\bar{x} - \Delta_{\bar{x}}\right) \le \sum_{i=1}^{N} x_i \le N\left(\bar{x} + \Delta_{\bar{x}}\right) \quad (4.39.)$$

Exemplul 4.3. Pentru problema din exemplul anterior, să se estimeze valoarea medie şi valoarea totală a creditelor din ultimul an, aplicând un sondaj aleatoriu simplu nerepetat.

Rezolvare:

Se calculează eroarea medie de reprezentativitate:

$$\sigma_{\bar{x}} \approx s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}} = \frac{530}{\sqrt{1000}}\sqrt{1-\frac{1000}{10000}} = 15{,}9 \text{ Euro}$$

Se determină eroarea limită:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}} = 1{,}96 \cdot 15{,}9 = 31{,}16 \text{ Euro}$$

Scanned with OKEN Scanner

Se determină intervalul de încredere pentru valoarea medie a creditelor:

$$8270 - 31{,}16 \le \mu \le 8270 + 31{,}16$$

$$8238{,}84 \le \mu \le 8301{,}16 \text{ Euro}$$

Dacă vrem să estimăm şi valoarea totală a creditelor, se înmulţeşte ultima dublă inegalitate cu 10.000, rezultând:

$$82.388.400 \le \sum_{i=1}^{N} x_i \le 83.011.600 \text{ Euro}$$

Se observă că în acest ultim caz, eroarea medie de reprezentativitate şi eroarea limită sunt mai mici decât în cazul sondajului repetat şi implicit lungimea intervalului de încredere va fi mai mică, deci estimarea va fi mai exactă (mai precisă).

4.6.2. Dimensionarea eşantionului în cazul sondajului simplu aleatoriu

Am văzut că între precizia şi siguranţa estimaţiei există o relaţie de inversă proporţionalitate şi că ambele pot fi influenţate de mărimea eşantionului. Deseori, se pune problema de a stabili volumul eşantionului pentru o anumită precizie (dată de o anumită eroare $\Delta_{\bar{x}}$) şi o anumită probabilitate de garantare a rezultatelor sondajului $100(1 - \alpha)\%$ (deci un anumit z).

Pentru ***sondajul simplu repetat***, dimensionarea eşantionului se deduce din relaţia de calcul a erorii medii:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sigma_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad (4.40.)$$

Prin prelucrări succesive, se obţine următoarea expresie a volumului eşantionului (n):

$$n = \frac{z_{\alpha/2}^2 s^2}{\Delta_{\bar{x}}^2} = \frac{s^2}{\sigma_{\bar{x}}^2} \qquad (4.41.)$$

Relaţia (4.41.) arată că volumul eşantionului este direct proporţional cu dispersia şi cu probabilitatea de garantare a rezultatelor sondajului şi invers proporţional cu eroarea limită. Aceasta înseamnă că dacă populaţia prezintă o variabilitate mai mare din punctul de vedere al caracteristicii urmărite, atunci va trebui să includem mai multe unităţi în eşantion. Dacă vrem să obţinem o eroare limită mai mică (celelalte elemente rămânând nemodificate), atunci va trebui să mărim volumul eşantionului.

Dispersia caracteristicii la nivelul eşantionului (s^2) este o estimaţie a dispersiei variabilei din întreaga colectivitate (σ^2), aceasta din urmă fiind, în general, necunoscută. Pentru dimensionarea eşantionului se poate folosi o valoare a dispersiei dintr-o cercetare anterioară prin sondaj.

Scanned with OKEN Scanner

Ştim că în cazul unei distribuţii cu tendinţă de normalitate aproximativ 99,73% dintre observaţii (adică aproape toţi termenii seriei) se regăsesc într-un interval aflat la o distanţă de cel mult 3 abateri medii pătratice de o parte şi de alta a mediei, adică în intervalul $\left(\bar{x}-3\sigma;\bar{x}+3\sigma\right)$, de lungime egală cu de 6 ori abaterea medie pătratică, prin urmare se poate aproxima amplitudinea variaţiei caracteristicii şi se poate determina că:

$$s \cong \frac{A_x}{6} \qquad (4.42.)$$

Cum n poate lua doar valori întregi, valoarea lui n, reieşită din calcul trebuie rotunjită la întregul imediat superior, pentru a fi siguri că eroarea limită este mai mică sau cel mult egală cu cea dorită (adică pentru a fi siguri că se respectă precizia dorită a estimaţiei).

Exemplul 4.4. Un cercetător doreşte să estimeze durata sejurului pentru cei 5000 de turişti care şi-au petrecut concediul într-o staţiune montană. Pentru aceasta, el a recurs la o cercetare parţială prin sondaj repetat. Dacă cercetătorul doreşte să garanteze rezultatele sondajului cu o probabilitate de 95,45% ($z = 2$) şi cu o eroare limită de 0,514 zile, să se determine câţi turişti ar trebui incluşi în eşantion, ştiind că, dintr-o cercetare anterioară, a rezultat o dispersie a duratei sejurului de 12,5 zile.

Rezolvare:

$$n = \frac{z_{\alpha/2}^2 s^2}{\Delta_{\bar{x}}^2} = \frac{4 \cdot 12{,}5}{(0{,}514)^2} = 189{,}25 \approx 190 \text{ turişti}$$

Pentru dimensionarea eşantionului în cazul ***sondajului simplu nerepetat***, se utilizează următoarea ecuaţie cu necunoscuta n:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{1-\frac{n}{N}} = z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} \qquad (4.43.)$$

Se fac următoarele prelucrări succesive:

$$\Delta_{\bar{x}}^2 = z_{\alpha/2}^2 \cdot \frac{s^2}{n}\left(1-\frac{n}{N}\right) = \frac{z_{\alpha/2}^2 s^2}{n} - \frac{z_{\alpha/2}^2 s^2}{N} \qquad (4.44.)$$

$$\frac{z_{\alpha/2}^2 s^2}{n} = \frac{z_{\alpha/2}^2 s^2}{N} + \Delta_{\bar{x}}^2 \qquad (4.45.)$$

$$n = \frac{z_{\alpha/2}^2 s^2}{\frac{z_{\alpha/2}^2 s^2}{N} + \Delta_{\bar{x}}^2} \qquad (4.46.)$$

Scanned with OKEN Scanner

Din relația (7) se observă că dacă volumul colectivității generale este foarte mare (adică $N \to \infty$), atunci $\frac{z_{\alpha/2}^2 s^2}{N} \to 0$ și deci relația devine:

$$n = \frac{z_{\alpha/2}^2 s^2}{\Delta_{\bar{x}}^2} \qquad (4.47.)$$

adică pentru N foarte mare, volumul eșantionului în cazul sondajului nerepetat se determină ca și în cazul sondajului repetat.

Exemplul 4.5. În exemplul anterior, să se determine câți turiști ar trebui incluși în eșantion dacă acesta a fost format prin selecție aleatorii nerepetată:

$$n = \frac{z_{\alpha/2}^2 s^2}{\frac{z_{\alpha/2}^2 s^2}{N} + \Delta_{\bar{x}}^2} = \frac{4 \cdot 12{,}5}{\frac{4 \cdot 12{,}5}{5000} + (0{,}514)^2} = \frac{50}{0{,}274196} = 182{,}35 \cong 183 \text{ turiști}$$

4.7. Sondajul aleatoriu simplu pentru o caracteristică alternativă

4.7.1. Determinarea erorilor și a intervalului de încredere pentru proporții

Dacă variabila urmărită arată îndeplinirea sau neîndeplinirea unei condiții, posedarea sau nu a unui anumit atribut, atunci este vorba de o caracteristică alternativă (w). În cazul unei variabile calitative alternative media îmbracă forma unei proporții și este definită ca ponderea cazurilor favorabile în numărul total de cazuri. Am văzut că această proporție este notată, la nivelul colectivității generale, cu π și la nivelul eșantionului cu p. Exemple de astfel de caracteristici alternative pot fi: opinia cetățenilor privind un nou act legislativ intrat în vigoare (pro/contra), starea civilă (căsătorit/necăsătorit), salariul (sub medie/peste medie), mediul de proveniență (urban/rural) etc.

Așadar, estimarea mediei înseamnă, în cazul unei variabile binare/alternative, estimarea proporției unităților care îndeplinesc o anumită condiție sau care posedă un anumit atribut (numărul cazurilor favorabile) în numărul total de cazuri. Parametrul la nivelul colectivității generale, care trebuie estimat, este proporția $\pi = \frac{M}{N}$, unde M este numărul de cazuri favorabile din populația totală, iar N este numărul total de unități din populația generală (volumul colectivității generale). Parametrul din populația generală va fi estimat cu ajutorul mediei de selecție $p = \frac{m}{n}$, unde m este numărul de unități din eșantion care sunt în stare

Scanned with OKEN Scanner

favorabilă, n este volumul eşantionului. Proporţiile de selecţie variază de la un eşantion la altul şi, pentru eşantioane de volum mare, această distribuţie de eşantionare este în concordanţă cu teorema limită centrală.

Conform acestei teoreme, dacă selectăm aleatoriu un eşantion de volum suficient de mare n, ($n > 30$) dintr-o populaţie totală de volum N, cu o pondere π a unităţilor aflate în stare favorabilă (sau care posedă un anumit atribut), atunci distribuţia de eşantionare a proporţiilor de selecţie p este aproximativ normal distribuită, de medie π. Spunem că p este un estimator nedeplasat al lui π.

Dispersia proporţiilor de selecţie va fi de n ori mai mică decât dispersia din colectivitatea generală:

$$\sigma_{\pi}^2 = \frac{\pi(1-\pi)}{n} \qquad (4.48.)$$

Cum dispersia din colectivitatea generală a caracteristicii alternative w nu este, de obicei, cunoscută (deoarece nici ponderea π nu este cunoscută), ea va fi estimată prin dispersia proporţiilor din eşantioane:

$$s_p^2 = \frac{p(1-p)}{n} \qquad (4.49.)$$

Abaterea medie pătratică a proporţiilor de selecţie va fi:

$$\sigma_p = s_p = \sqrt{\frac{p(1-p)}{n}} \text{ pentru selecţia repetată} \qquad (4.50.)$$

$$\sigma_p = s_p = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \text{ pentru selecţia nerepetată} \qquad (4.51.)$$

şi se numeşte ***eroare medie de reprezentativitate***.

Intervalul de încredere, despre care se poate afirma cu un nivel de încredere $(1 - \alpha)$ că include proporţia la nivelul colectivităţii generale este construit cu ajutorul ***erorii limită***:

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p \qquad (4.52.)$$

adică:

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \text{ în cazul selecţiilor repetate} \qquad (4.53.)$$

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \text{ în cazul selecţiilor nerepetate} \qquad (4.54.)$$

Intervalul de încredere va fi de forma:

$$p - \Delta_p \leq \pi \leq p + \Delta_p \qquad (4.55.)$$

Scanned with OKEN Scanner

adică:

$$p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$ pentru selecția repetată (4.56.)

$$p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \le \pi \le p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}$$ pentru selecția nerepetată. (4.57.)

Dacă vrem să estimăm numărul de unități statistice din colectivitatea generală care sunt în stare favorabilă (M), intervalul de încredere va fi:

$$N(p - \Delta_p) \le M \le N(p + \Delta_p) \quad (4.58.)$$

adică:

$$N\left(p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right) \le M \le N\left(p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right)$$ pentru selecția repetată. (4.59.)

$$N\left(p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}\right) \le M \le N\left(p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}\right)$$ pentru selecția nerepetată. (4.60.)

Exemplul 4.6. Pentru estimarea ponderii turiştilor nemulţumiţi de serviciile de care au beneficiat din partea unei mari agenţii de turism, s-au selectat aleatoriu 500 din cei 5000 de clienţi ai agenţiei din ultimul an. În urma sondajului, a rezultat că 75 dintre clienţi nu au fost mulţumiţi de serviciile oferite. Să se estimeze, pentru o probabilitate de 95% ($z = 1{,}96$), ponderea clienţilor nemulţumiţi de serviciile agenţiei, în ipoteza selecţiei simple aleatorii repetate şi nerepetate. Comparaţi rezultatele obţinute.

Rezolvare:

a) Sondaj aleatoriu simplu repetat:

$$p = \frac{m}{n} = \frac{75}{500} = \frac{15}{100} = 0{,}15$$

$$s_w^2 = p(1-p) = 0{,}15 \cdot (1 - 0{,}15) = 0{,}1275$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}1275}{500}} = 0{,}016 \ (1{,}6\%)$$

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p = 1{,}96 \cdot 0{,}016 = 0{,}031 \ (3{,}1\%)$$

$$0{,}15 - 0{,}031 \le \pi \le 0{,}15 + 0{,}031$$

$$0{,}119 \le \pi \le 0{,}181$$

Scanned with OKEN Scanner

Aşadar, cu o probabilitate de 95%, putem afirma că ponderea clienţilor nemulţumiţi de serviciile oferite de agenţie este între 11,9% şi 18,1%.

b) Sondaj aleatoriu simplu nerepetat:

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}1275}{500}\cdot\left(1-\frac{500}{5000}\right)} = 0{,}015\,(1{,}5\%)$$

$$\Delta_p = z_{\alpha/2}\cdot\sigma_p = 1{,}96\cdot 0{,}015 = 0{,}029\ (2{,}9\%)$$

$$0{,}15 - 0{,}029 \le \pi \le 0{,}15 + 0{,}029$$

$$0{,}121 \le \pi \le 0{,}179$$

Aşadar, cu o probabilitate de 95%, se poate afirma că ponderea clienţilor nemulţumiţi de serviciile oferite de agenţie este între 12,1% şi 17,9%.

În cazul sondajului nerepetat s-a obţinut o eroare mai mică decât în cazul sondajului repetat, deci un interval de încredere mai mic, care a dus la o estimare mai precisă (exactă).

4.7.2. Dimensionarea eşantionului în cazul unei variabile alternative

Volumul eşantionului în cazul variabilei alternative se determină în acelaşi mod ca şi în cazul aplicării sondajului pentru o caracteristică nealternativă.

Astfel, în cazul selecţiei repetate, volumul eşantionului se determină din formula erorii limită:

$$\Delta_p = z_{\alpha/2}\cdot\sqrt{\frac{p(1-p)}{n}} \qquad (4.61.)$$

din care se scrie expresia lui n:

$$n = \frac{z^2_{\alpha/2}\,p(1-p)}{\Delta^2_p} \qquad (4.62.)$$

Similar, în cazul selecţiei nerepetate, obţinem:

$$\Delta_p = z^2_{\alpha/2}\cdot\sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \qquad (4.63.)$$

din care se scrie expresia lui n:

$$n = \frac{z^2_{\alpha/2}\,p(1-p)}{\frac{z^2_{\alpha/2}\,p(1-p)}{N} + \Delta^2_p} \qquad (4.64.)$$

Aşadar, se observă că şi în cazul sondajului aplicat pentru caracteristică alternativă, volumul eşantionului depinde de următorii factori:

- eroarea limită maxim admisibilă (Δ_p);

Scanned with OKEN Scanner

- nivelul de încredere sau probabilitate cu care se garantează rezultatele: $100(1-\alpha)\,p$, respectiv z;
- dispersia proporţiilor la nivelul colectivităţii generale: $\pi(1-\pi)$, respectiv dispersia proporţiilor la nivelul eşantionului: $p(1-p)$;
- volumul colectivităţii generale (în cazul selecţiei nerepetate).

Ca şi în cazul variabilei nealternative, dispersia caracteristicii alternative se poate cunoaşte dintr-o cercetare anterioară sau se poate înlocui cu valoarea maximă a dispersiei în cazul variabilelor binare: 0,25, obţinută atunci când ponderea cazurilor favorabile este egală cu ponderea cazurilor nefavorabile.

Exemplul 4.7. Pentru exemplul anterior, câţi turişti ar trebui incluşi în eşantion dacă dorim să obţinem o eroare limită mai mică cu 20% decât în cazul anterior? Aplicaţi selecţia aleatorie simplă repetată şi nerepetată. Celelalte elemente rămân nemodificate.

Rezolvare:

a) pentru sondaj repetat:

$$\Delta'_p = \frac{80}{100}\cdot\Delta_p = 0{,}8\cdot 0{,}031 = 0{,}025$$

$$n' = \frac{z^2_{\alpha/2}\,p(1-p)}{\Delta'^2_p} = \frac{1{,}96^2\cdot 0{,}1275}{(0{,}025)^2} = \frac{0{,}4898}{0{,}000625} = 783{,}68 \cong 784 \text{ clienţi}$$

b) pentru sondaj nerepetat:

$$\Delta'_p = \frac{80}{100}\cdot\Delta_p = 0{,}8\cdot 0{,}029 = 0{,}0232$$

$$n' = \frac{z^2_{\alpha/2}\,p(1-p)}{\frac{z^2_{\alpha/2}\,p(1-p)}{N} + \Delta'^2_p} = \frac{1{,}96^2\cdot 0{,}1275}{\frac{1{,}96^2\cdot 0{,}1275}{5000} + (0{,}025)^2} = \frac{0{,}4898}{0{,}0006362} = 769{,}88 \cong 770 \text{ clienţi}$$

4.8. Determinarea erorilor şi a intervalului de încredere în cazul eşantioanelor de volum redus

Metodele pe care le-am prezentat până acum pentru determinarea erorilor şi a intervalului de încredere pentru parametrul (media) colectivităţii generale sunt valabile atunci când populaţia din care s-a selectat eşantionul este normal distribuită, sau când eşantionul este de volum mare ($n > 30$). Utilizarea abaterilor normal-normate z şi a testului cu acelaşi nume are, însă, o serie de limite:

- nu cunoaştem valorile parametrilor din colectivitatea generală, de exemplu abaterea medie pătratică din colectivitatea totală (σ), caz în care vom

Scanned with OKEN Scanner

estima parametrii necunoscuţi prin aplicarea sondajului statistic, pe baza informaţiilor oferite de eşantion;

- dacă lucrăm cu eşantioane de volum redus, abaterea medie pătratică a caracteristicii urmărite, la nivelul eşantionului (s), este un estimator inexact al abaterii medii pătratice din colectivitatea generală (σ). În general, pentru eşantioane ce includ mai puţin de 30 de unităţi statistice, în peste jumătate din cazuri, abaterea standard la nivelul eşantionului subestimează abaterea standard a variabilei studiate la nivelul colectivităţii generale, iar distribuţia normală nu mai descrie cu acurateţe distribuţia de eşantionare a mediilor de selecţie.

În acelaşi mod se poate afirma că dispersia $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ (pentru o serie simplă), sau $s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i}$ (pentru o serie de frecvenţe) a eşantioanelor este un estimator deplasat al dispersiei σ^2 din colectivitatea generală. Pentru a determina, totuşi, un estimator nedeplasat, mai bun, al dispersiei σ^2 se foloseşte pentru calcularea dispersiei eşantionului următoarea formulă de calcul:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad \text{(pentru serii simple)} \qquad (4.65.)$$

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{r} n_i - 1} \quad \text{(pentru serii de frecvenţe)} \qquad (4.66.)$$

Dacă eşantionul este de volum redus, distribuţia de eşantionare a mediei de selecţie poate fi descrisă printr-o distribuţie Student (t); între distribuţia Student şi distribuţia normală există o serie de similarităţi, dar şi de diferenţe:

a) similarităţi:

- ambele distribuţii sunt distribuţii de eşantionare; aşa cum reiese din Teorema limită centrală, distribuţia de eşantionare a mediilor de selecţie, pentru eşantioane generate în mod aleatoriu din orice populaţie totală – este o distribuţie de tip normal, dacă eşantioanele formate sunt de volum mare. Distribuţia Student este tot o distribuţie de eşantio-

Scanned with OKEN Scanner

nare, care descrie distribuția de eșantionare preconizată, previzionată pentru eșantioane de un anumit volum specificat;

- ambele distribuții sunt simetrice, având formă de clopot;
- ambele distribuții au medie zero, ceea ce indică modul de folosire a acestora. În cazul unei distribuții normale, se utilizează testul z pentru a determina probabilitatea necesară obținerii unei anumite valori a caracteristicii și cu cât diferența (abaterea) între medie și valoarea observată este mai mare, cu atât valoarea absolută a testului este mai mare. Același lucru se poate spune și în cazul distribuției z, ambele distribuții putându-se folosi pentru a determina probabilitatea cu care se obține o anumită medie de selecție, pentru o anumită abatere relativă dată a acesteia față de media colectivității totale.

b) diferențe:

- există câte o distribuție t separată pentru fiecare volum (dimensiune) a eșantionului, fiecare distribuție Student reprezentând numărul de grade de libertate pentru eșantioanele respective. De exemplu, distribuția t pentru un grad de libertate ($ngl = 1$) reprezintă distribuția de eșantionare atunci când avem două observații și, similar, pentru 7 grade de libertate ($ngl = 7$) distribuția Student reprezintă distribuția de eșantionare în care media și abaterea standard au fost obținute pe baza a 8 observații;
- pe măsură ce mărimea eșantionului crește, distribuțiile Student se apropie de distribuția normală, diferențe mai mari între cele două distribuții apărând pentru volume reduse ale eșantioanelor.

Dacă formăm în mod aleatoriu repetat eșantioane de volume n dintr-o populație totală de volum N, în locul abaterilor normal normate $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$, se determină acum:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad (4.67.)$$

care urmează o repartiție Student, cu următoarele caracteristici:

- este în formă de clopot;
- este simetrică față de zero;
- este mai larg dispersată față de distribuția normală;
- forma ei depinde de volumul eșantionului (altfel spus, dacă n = volumul eșantionului, distribuția Student prezintă $n - 1$ grade de libertate).

În cazul în care dispersia din colectivitatea generală (σ^2) nu este cunoscută, estimarea parametrului μ al colectivității generale se face prin intervalul de încredere de forma:

Scanned with OKEN Scanner

$$\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \qquad (4.68.)$$

unde α reprezintă nivelul de semnificație.

Valorile statisticii $t_{\alpha/2,n-1}$ se regăsesc în tabelele Student pentru un număr de grade de libertate $n - 1$ (de la 1 la 29). Pentru un număr de grade de libertate de peste 30 ($ngl > 30$) și dacă dispersia din colectivitatea generală nu este cunoscută, $t_{\alpha/2}$ este aproximativ egal cu $z_{\alpha/2}$.

Exemplul 4.8. Managerul unei societăți de tranzacții imobiliare află că valoarea medie de piață pentru 20 de case dintr-un anumit cartier al unui oraș mare este de 155.000 de Euro, cu o abatere medie pătratică de 17.200 de Euro. Ajutați-l pe manager să estimeze valoarea medie de piață a unei case din întregul oraș, cu o probabilitate de 90% (presupunem că valoarea de piață a caselor urmează o distribuție normală).

Rezolvare:

$n = 20$

$\bar{x} = 155000$ Euro

$s = 17200$ Euro

$\alpha = 1 - 0{,}9 = 0{,}1$

$$\frac{\alpha}{2} = 0{,}05$$

$$t_{\alpha/2,n-1} = t_{0{,}05;19} = 2{,}093$$

Intervalul de încredere pentru valoarea medie de piață a unei case din întregul oraș va fi:

$$\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

adică:

$$155000 - 2{,}093 \cdot \frac{17200}{\sqrt{20}} \leq \mu \leq 155000 + 2{,}093 \cdot \frac{17200}{\sqrt{20}}$$

$(146950 \leq \mu \leq 163050)$ Euro

Scanned with OKEN Scanner

4.9. Sondajul aleatoriu tipic (stratificat)

Eşantionarea stratificată se realizează atunci când există un grad mare de dispersare a unităţilor colectivităţii totale, din punctul de vedere al caracteristicii studiate şi constă în formarea unor straturi omogene (care au un nivel redus de variaţie). În acest fel, se poate reduce variabilitatea în cadrul populaţiei, lucru ce va avea drept efect micşorarea valorii dispersiei şi, întrucât eroarea de sondaj depinde direct proporţional de mărimea dispersiei, se poate spune că eşantionarea stratificată va duce la o diminuare a erorii de selecţie şi deci la o creştere a exactităţii estimaţiei. Din cele prezentate anterior în acest capitol, am văzut că din fiecare strat astfel format se poate extrage aleatoriu simplu câte un subeşantion, inferenţa efectuându-se atât la nivelul fiecărui strat, cât şi la nivelul întregii populaţii.

4.9.1. Noţiuni specifice pereche utilizate în cadrul sondajului tipic

- Volumul colectivităţii

Colectivitatea totală, de volum N, este împărţită în r straturi (grupe), de volum Ni, $i=\overline{1,r}$.

$$N_1 + N_2 + ... + N_r = \sum_{i=1}^{r} N_i = N \tag{4.69.}$$

Din fiecare strat i al populaţiei totale se extrage aleatoriu simplu câte un subeşantion, de volum n_i, $i=\overline{1,r}$. În acest fel, cele r subeşantioane compun un eşantion de volum n.

$$n_1 + n_2 + ... + n_r = \sum_{i=1}^{r} n_i = n \tag{4.70.}$$

A) Pentru o caracteristică nealternativă, avem:

- **Media**

Media stratului i din colectivitatea generală se determină:

$$\mu_i = \frac{\sum_{j=1}^{Ni} x_{ij}}{N_i} \tag{4.71.}$$

Media la nivelul colectivităţii generale se va forma din mediile tuturor straturilor:

$$\mu = \frac{\sum_{i=1}^{r} \mu_i N_i}{\sum_{i=1}^{r} N_i} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{Ni} x_{ij}}{N} \tag{4.72.}$$

Scanned with OKEN Scanner

Media subeşantionului i din eşantion se va determina:

$$\bar{x}_i = \frac{\sum_{j=1}^{ni} x_{ij}}{n_i} \qquad (4.73.)$$

iar media la nivelul eşantionului poate fi o medie aritmetică ponderată a medilor subeşantioanelor:

$$\bar{x} = \frac{\sum_{i=1}^{r} \bar{x}_i n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{ni} x_{ij}}{n} \qquad (4.74.)$$

- **Dispersia**

Pentru aflarea dispersiei pe întreaga colectivitate, scriem regula de adunare a dispersiilor:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2 \qquad (4.75.)$$

unde:

σ^2 este dispersia totală la nivelul colectivităţii generale;

$\bar{\sigma}^2$ este media dispersiilor straturilor în colectivitatea generală;

δ^2 este dispersia dintre straturile colectivităţii totale.

În cazul sondajului tipic, se consideră că dispersia dintre straturi este nulă ($\delta^2 = 0$), întrucât presupunem că populaţia totală era de la început împărţită pe cele r straturi independente, astfel că dispersia totală rămâne egală cu media dispersiilor straturilor:

$$\sigma^2 = \bar{\sigma}^2 \qquad (4.76.)$$

La nivelul eşantionului, media dipersiilor straturilor va fi aproximată de indicatorul pereche aferent colectivităţii de selecţie:

$$\bar{\sigma}^2 \approx \bar{s}^2 = \frac{\sum_{i=1}^{r} s_i^2 n_i}{\sum_{i=1}^{r} n_i} \qquad (4.77.)$$

unde s_i^2 reprezintă dispersia variabilei urmărite în subeşantionul i.

$$n_1 + n_2 + \ldots + n_r = \sum_{i=1}^{r} n_i = n \qquad (4.78.)$$

Scanned with OKEN Scanner

B) Pentru o caracteristică alternativă, formulele devin:

- **Media**

Media stratului *i* din colectivitatea generală se determină:

$$\pi_i = \frac{M_i}{N_i} \tag{4.79.}$$

Media totală a colectivităţii va fi:

$$\pi = \frac{M}{N} = \frac{\sum_{i=1}^{r} M_i}{\sum_{i=1}^{r} N_i} = \frac{\sum_{i=1}^{r} \pi_i N_i}{\sum_{i=1}^{r} N_i} \tag{4.80.}$$

Media subeşantionului *i* din eşantion se determină cu relaţia:

$$p_i = \frac{m_i}{n_i} \tag{4.81.}$$

Media eşantionului este dată de:

$$p = \frac{m}{n} = \frac{\sum_{i=1}^{r} p_i n_i}{\sum_{i=1}^{r} n_i} \tag{4.82.}$$

- **Dispersia**

Media dispersiilor grupelor din colectivitatea generală este:

$$\overline{\pi(1-\pi)} = \frac{\sum_{i=1}^{r} \pi_i (1-\pi_i) N_i}{\sum_{i=1}^{r} N_i} \tag{4.83.}$$

iar media dispersiilor subeşantioanelor:

$$\overline{p(1-p)} = \frac{\sum_{i=1}^{r} p_i (1-p_i) n_i}{\sum_{i=1}^{r} n_i} \tag{4.84.}$$

Scanned with OKEN Scanner

4.9.2. Determinarea erorilor, a intervalului de încredere şi a volumului eşantionului în cazul sondajului aleatoriu tipic

A) În cazul unei variabile nealternative:

Pentru o ***eşantionare stratificată repetată***, relaţiile de calcul sunt:

Eroarea medie de reprezentativitate va fi:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\bar{\sigma}^2}{n}} \approx \sqrt{\frac{\bar{s}^2}{n}} \qquad (4.85.)$$

Cum $\bar{\sigma}^2 < \sigma^2$, rezultă că eroarea medie de reprezentativitate în cazul sondajului stratificat este mai mică decât eroarea medie de reprezentativitate în cazul sondajului aleatoriu simplu.

Eroarea limită maxim admisibilă va fi:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sigma_{\bar{x}} \qquad (4.86.)$$

unde $z_{\alpha/2}$ reprezintă argumentul funcţiei de probabilitate, pentru o probabilitate de garantare a rezultatelor sondajului de $100(1 - \alpha)\%$, α fiind pragul de semnificaţie.

Intervalul de încredere pentru estimarea parametrului μ al colectivităţii generale va fi:

$$\bar{x} - \Delta_{\bar{x}} \le \mu \le \bar{x} + \Delta_{\bar{x}} \qquad (4.87.)$$

adică:

$$\bar{x} - z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}} \le \mu \le \bar{x} + z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}} \qquad (4.88.)$$

Volumul eşantionului se determină rezolvând ecuaţia dată de formula erorii limită, ce are ca necunoscută volumul n al eşantionului:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}} \qquad (4.89.)$$

de unde:

$$n = \frac{z_{\alpha/2}^2 \bar{s}^2}{\Delta_{\bar{x}}^2} \qquad (4.90.)$$

În cazul ***selecţiei aleatorii stratificate nerepetate***, se introduce în relaţia de calcul a erorii medii de reprezentativitate coeficientul de corecţie:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{\bar{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)} \approx \sqrt{\frac{\bar{s}^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (4.91.)$$

Scanned with OKEN Scanner

Eroarea limită devine:

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (4.92.)$$

Intervalul de încredere prin care se estimează parametrul μ al colectivității generale, pentru probabilitatea de $100(1-\alpha)\%$, va fi:

$$\bar{x} - \Delta_{\bar{x}} \leq \mu \leq \bar{x} + \Delta_{\bar{x}} \qquad (4.93.)$$

adică:

$$\bar{x} - z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}\left(1 - \frac{n}{N}\right)} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \sqrt{\frac{\bar{s}^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (4.94.)$$

Volumul eşantionului va fi dat de relaţia:

$$n = \frac{z_{\alpha/2}^2 \bar{s}^2}{\frac{z_{\alpha/2}^2 \bar{s}^2}{N} + \Delta_{\bar{x}}^2} \qquad (4.95.)$$

Exemplul 4.9. O companie producătoare de software specializat a creat trei programe pentru a facilita studenţilor aprofundarea ştiinţelor economice. Pentru testarea eficacităţii acestor programe, au fost alcătuite trei grupuri de studenţi, fiecare grup fiind instruit cu ajutorul câte unuia dintre programe. După două săptămâni de instruire, studenţii au primit un test de verificare a cunoştinţelor la ştiinţele economice. Rezultatele sistematizate se prezintă astfel:

Tabelul 4.4.

Grupe de studenţi după programul tutorial	Număr de studenţi (n_i)	Punctajul mediu ($\bar{x}_i$)	Abaterea medie pătratică a punctajului (s_i)	Ponderea studenţilor care au obţinut peste 90 de puncte (p_i)
Programul 1	20	88	3,5	12
Programul 2	50	60	6,2	10
Programul 3	30	92	5,8	15

Ştiind că cei 100 de studenţi formează un eşantion de 10% din populaţia generală, se cere să se estimeze punctajul mediu al unui student din colectivitatea generală, cu o probabilitate de 99,73% ($z = 3$), aplicând un sondaj tipic nerepetat.

Rezolvare:

Sondaj stratificat nerepetat:

$$\bar{x} = \frac{\sum \bar{x}_i n_i}{\sum n_i} = \frac{88 \cdot 20 + 60 \cdot 50 + 92 \cdot 30}{100} = 75{,}2 \text{ puncte}$$

Scanned with OKEN Scanner

$$s^2 = \bar{s}^2 + d^2$$

$$\bar{s}^2 = \frac{\sum s_i^2 n_i}{\sum n_i} = \frac{3{,}5^2 \cdot 20 + 6{,}2^2 \cdot 50 + 5{,}8^2 \cdot 30}{100} = 31{,}76$$

$$\sigma_{\bar{x}} \approx s_{\bar{x}} = \sqrt{\frac{\bar{s}_x^2}{n}\left(1 - \frac{n}{N}\right)} = 0{,}53 \text{ puncte}$$

$$\Delta_{\bar{x}} = z_{\alpha/2} \cdot \sigma_{\bar{x}} = 3 \cdot 0{,}53 = 1{,}6 \text{ puncte}$$

$$\bar{x} - \Delta_{\bar{x}} \le \mu \le \bar{x} + \Delta_{\bar{x}}$$

$$75{,}2 - 1{,}6 \le \mu \le 75{,}2 + 1{,}6 \text{ puncte}$$

$$73{,}6 \le \mu \le 76{,}8 \text{ puncte}$$

Odată determinat volumul eşantionului, în cazul eşantionării stratificate, apare o problemă suplimentară faţă de selecţia simplă şi anume: cum se va repartiza volumul determinat al eşantionului pe subeşantioane? Pentru aceasta, există două modalităţi:

i) ***repartizarea proporţională*** a volumului eşantionului pe straturi (subeşantioane) presupune că în fiecare subeşantion se va repartiza o aceeaşi proporţie de unităţi din volumul eşantionului, ca şi în colectivitatea generală, adică fracţiile de sondaj sunt egale pentru toate straturile:

$$\frac{n_i}{n} = \frac{N_i}{N} \Leftrightarrow n_i = \frac{N_i}{N} \cdot n \qquad (4.96.)$$

ii) ***repartizarea optimă*** a volumului eşantionului pe straturi ţine cont, în plus, şi de variaţia din cadrul stratului respectiv: cu cât variaţia din cadrul unui strat este mai mare, cu atât se va repartiza în acel strat o proporţie mai mare de unităţi din volumul eşantionului, după relaţia:

$$n_i = \frac{s_i N_i}{\sum_{i=1}^{r} s_i N_i} \cdot n \qquad (4.97.)$$

Exemplul 4.10. În problema anterioară, aflaţi câţi studenţi ar trebui incluşi în eşantion, dacă s-ar organiza un nou sondaj tipic, cu o eroare limită mai mică de 1,5 ori? Repartizaţi-i proporţional şi optim pe subeşantioane.

Rezolvare:

$$\Delta'_{\bar{x}} = \frac{\Delta_{\bar{x}}}{1{,}5} = 1{,}067 \text{ puncte}$$

Scanned with OKEN Scanner

$$n' = \frac{z_{\alpha/2}^2 \bar{s}^{-2}}{\Delta_{\bar{x}}^{'2} + \frac{z_{\alpha/2}^2 \bar{s}^{-2}}{N}} = 200{,}7 \approx 201 \text{ studenţi}$$

Repartizarea noului eşantion pe subeşantioane este prezentată în tabelul următor: prin procedeul proporţional şi prin procedeul optim.

Tabelul 4.5.

Grupe de studenţi după programul tuturial	N_i	$\frac{N_i}{\Sigma N_i}$	$n' \cdot \frac{N_i}{\Sigma N_i}$	$s_i N_i$	$\frac{s_i N_i}{\sum s_i N_i}$	$n' \cdot \frac{s_i N_i}{\sum s_i N_i}$
Programul 1	200	0,2	40	700	0,13	26
Programul 2	500	0,5	101	3100	0,56	113
Programul 3	300	0,3	60	1740	0,31	62
Total	1000	1,0	201	5540	1,00	201

B) În cazul unei variabile alternative:

Eroarea medie de reprezentativitate va fi:

$$\sigma_p \approx s_p = \sqrt{\frac{p(1-p)}{n}} \text{ (pentru sondaj repetat)} \qquad (4.98.)$$

$$\sigma_p \approx s_p = \sqrt{\frac{p(1-p)}{n}\left(1 - \frac{n}{N}\right)} \text{ (pentru sondaj nerepetat)} \qquad (4.99.)$$

Eroarea limită va fi dată de relaţia:

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p \qquad (4.100.)$$

Intervalul de încredere se determină astfel:

$$p - \Delta_p \leq \pi \leq p + \Delta_p \qquad (4.101.)$$

Volumul eşantionului se calculează cu relaţiile:

$$n = \frac{z_{\alpha/2}^2 \cdot \overline{p(1-p)}}{\Delta_p^2} \text{ (pentru selecţie repetată)} \qquad (4.102.)$$

$$n = \frac{z_{\alpha/2}^2 \cdot \overline{p(1-p)}}{\frac{z_{\alpha/2}^2 \cdot \overline{p(1-p)}}{N} + \Delta_p^2} \text{ (pentru selecţie nerepetată)} \qquad (4.103.)$$

Exemplul 4.11. Pentru problema anterioară, să se estimeze ponderea studenţilor din populaţia totală care au obţinut la test mai mult de 90 de puncte.

Scanned with OKEN Scanner

Rezolvare:

$$p = \frac{\sum p_i n_i}{\sum n_i} = \frac{0{,}12 \cdot 20 + 0{,}1 \cdot 50 + 0{,}15 \cdot 30}{100} = 0{,}12$$

$$p_1(1-p_1) = 0{,}12 \cdot 0{,}88 = 0{,}1056$$

$$p_2(1-p_2) = 0{,}1 \cdot 0{,}9 = 0{,}09$$

$$p_3(1-p_3) = 0{,}15 \cdot 0{,}85 = 0{,}1275$$

$$\overline{p(1-p)} = \frac{\sum p_i(1-p_i)n_i}{\sum n_i} = \frac{0{,}1056 \cdot 20 + 0{,}09 \cdot 50 + 0{,}1275 \cdot 30}{100} = 0{,}104$$

$$\sigma_p = \sqrt{\frac{\overline{p(1-p)}}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}104}{100} \cdot 0{,}9} = 0{,}03 \quad (3{,}0\%)$$

$$\Delta_p = z_{\alpha/2} \cdot \sigma_p = 3 \cdot 0{,}03 = 0{,}09 \quad (9\%)$$

$$p - \Delta_p \le \pi \le p + \Delta_p$$

$$0{,}12 - 0{,}09 \le \pi \le 0{,}12 + 0{,}09$$

$$0{,}03 \le \pi \le 0{,}21$$

(3%) (21%)

4.10. Elemente de testare a ipotezelor statistice

Dacă în paragrafele anterioare am prezentat modalitatea în care se estimează diferiți parametri ai colectivității generale, în prezentul paragraf ne vom ocupa de al doilea aspect al inferenței statistice – testarea ipotezelor, al cărui scop este de a stabili existența unor informații, dovezi suficiente pentru a concluziona dacă o ipoteză sau o afirmație referitoare la un parametru este adevărată. Acest tip de inferență este absolut necesar în fundamentarea procesului decizional, pentru optimizarea conducerii fenomenelor economico-sociale.

Exemplul 4.12. Managerul unei companii producătoare de băuturi răcoritoare dorește să verifice efectul ultimei campanii publicitare a companiei asupra cotei de piață deținute. În acest scop, el organizează o cercetare selectivă în care înregistrează cota de piață deținută în 12 luni înainte și după lansarea și promovarea ultimei campanii publicitare. În acest caz, trebuie verificată ipoteza existenței unei diferențe semnificative între cota medie de piață de dinaintea campaniei publicitare și cea de după lansarea ei.

Scanned with OKEN Scanner

Exemplul 4.13. Managerul unui post de televiziune X este interesat în determinarea audienței unor programe de știri promovate de postul său. În acest scop, el comandă o cercetare prin sondaj pe un anumit număr de gospodării, înregistrându-se răspunsul la întrebarea: „Urmăriți știrile de la ora 19:00 ale postului de televiziune X? Cu ajutorul testelor statistice de verificare a ipotezelor se va estima care este proporția gospodăriilor care urmăresc programul de știri de la ora 19:00 al postului respectiv de televiziune.

Există două tipuri de ipoteze: ipoteze conceptuale și ipoteze statistice.

O ***ipoteză conceptuală*** este o afirmație generală referitoare la relația dintre variabile independente și variabile dependente, sau referitoare la nivelul unei observații.

O ***ipoteză statistică*** este o afirmație matematică ce descrie relația dintre dintre doi sau mai mulți parametri ai colectivității totale, căreia i se poate demonstra valoarea de adevăr sau de fals.

Testele statistice care verifică valoarea unui parametru al colectivității totale se numesc ***teste parametrice***. Un astfel de test are 4 componente:

a) ipoteza nulă;
b) ipoteza alternativă;
c) testul statistic propriu-zis;
d) adoptarea deciziei.

a) ***Ipoteza nulă***

Ipoteza nulă este o afirmație în care se specifică valori ipotetice pentru unul sau mai mulți parametri ai populației generale. De obicei, în formularea ipotezei nule (notată cu H_0) se utilizează expresia „nu există diferențe".

Exemplul 4.14. În dorința de a testa dacă numărul mediu de defecțiuni zilnice ale unei linii de producție este egal cu 3, ipoteza nulă va fi:

H_0: $\mu = 3$

Exemplul 4.15. În dorința de a stabili dacă proporția rebuturilor dintr-un lot de produse fabricat este de 2%, vom formula ipoteza nulă astfel:

H_0: $\pi = 0{,}02$

b) ***Ipoteza alternativă***

Ipoteza alternativă (notată H_1) este o afirmație despre același parametru al populației ca și cel la care se referă ipoteza nulă. Ipoteza alternativă afirmă, de obicei, că parametrul colectivității generale are o valoare diferită față de cea specificată în ipoteza nulă. Respingerea ipotezei nule implică automat acceptarea ipotezei alternative. Dacă, de regulă, ipoteza nulă prevede inexistența diferențelor (condiții de identitate absolută, de egalitate), ipoteza alternativă susține contrariul.

Scanned with OKEN Scanner

Ipotezele alternative pot fi, la rândul lor, de două tipuri:

b_1) ***Ipoteze direcționale*** – reprezintă ipoteze alternative care stabilesc direcția în care parametrul populației totale diferă de valoarea specificată în ipoteza nulă (sub sau peste, mai mare sau mai mic). Acest tip de ipoteze corespunde testelor unidirecționale (testul unilateral stâng sau drept).

Exemplul 4.16. Pentru exemplul 4.14., ipoteza alternativă direcțională poate fi:

H_1: $\mu > 3$ sau H_1: $\mu < 3$ (test unilateral dreapta sau stânga)

Pentru exemplul 4.15., ipoteza alternativă direcțională ar putea fi:

H_1: $\pi > 0{,}02$ sau H_1: $\pi < 0{,}02$ (test unilateral dreapta sau stânga)

b_2) ***Ipoteze nondirecționale*** – reprezintă ipoteze alternative care specifică existența unei diferențe între valoarea parametrului populației generale și valoarea specificată în ipoteza nulă, dar fără a arăta direcția acestei diferențe. Ipotezele alternative nondirecționale sunt asociate testelor bilaterale (parametrul populației generale este ori mai mare, ori mai mic decât valoarea specificată în ipoteza nulă).

Exemplul 4.17. Pentru exemplul (4.14.), ipoteza alternativă nondirecțională poate fi:

H_1: $\mu \neq 3$ (adică $\mu > 3$ sau $\mu < 3$)

Pentru exemplul (4.15.), ipoteza alternativă nondirecțională va fi:

H_1: $\pi \neq 0{,}02$ (adică $\pi > 0{,}02$ sau $\pi < 0{,}02$)

După formularea celor două ipoteze se consideră în continuare că ipoteza nulă este adevărată și asta până când demonstrăm că nu este așa.

În urma aplicării testului de verificare a ipotezelor statistice, se pot lua următoarele decizii:

- decizia în concordanță cu afirmația ipotezei nule (spunem deci, că nu se respinge ipoteza nulă H_0);
- decizia opusă afirmației din ipoteza nulă (spunem deci, că respingem ipoteza nulă H_0).

c) ***Testul statistic***

Testul statistic reprezintă un criteriu pe baza căruia luăm decizia de a respinge sau a nu respinge ipoteza nulă. Testul statistic este o variabilă aleatorie a cărei valoare va fi utilizată pentru adoptarea uneia sau alteia dintre decizii.

Testul statistic se bazează pe un estimator punctual al parametrului ce va fi testat.

Exemplul 4.18. Pentru a testa media unei populații totale, dacă este cunoscută abaterea medie pătratică, atunci vom utiliza testul statistic:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad (4.104.)$$

Scanned with OKEN Scanner

Pentru a testa dacă proporţia unităţilor dintr-o populaţie totală, care îndeplinesc o anumită condiţie (caracteristică alternativă), testul va fi de forma:

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} \qquad (4.105.)$$

d) ***Regula de decizie***

Regula de decizie defineşte intervalul de valori ale unui test statistic care duce la respingerea ipotezei nule şi care este denumit ***regiune de respingere*** sau ***regiune critică***. Aşadar, dacă valoarea testului statistic se găseşte în regiunea critică, atunci vom respinge ipoteza nulă H_0. Dacă valoarea testului statistic nu se regăseşte în regiunea critică, nu avem suficiente argumente pentru a respinge ipoteza nulă H_0. Intervalul de valori ale testului statistic care nu sunt în regiunea critică se numeşte ***regiune noncritică*** sau ***regiune de acceptare*** sau ***interval de încredere***.

Întrucât concluziile la care ajungem se bazează, de cele mai multe ori, pe date ce provin din eşantioane (date parţiale), există implicit şi posibilitatea de a comite unele erori. Astfel, ipoteza nulă poate fi adevărată sau falsă şi, pe baza rezultatelor testului statistic, obţinute cu informaţii înregistrate de la unităţile statistice ale eşantioanelor, vom decide dacă acceptăm sau dacă respingem ipoteza nulă (Tabelul 4.6.). Pot apărea, de aceea, următoarele decizii corecte:

- să nu respingem ipoteza nulă (să o acceptăm) atunci când ea este adevărată;
- să respingem ipoteza nulă atunci când ea este falsă.

Totodată, există şi două decizii incorecte pe care le putem adopta pe baza valorii testului statistic:

- să respingem ipoteza nulă atunci când ea este adevărată (această eroare se numeşte „eroare de tipul I“);
- să nu respingem ipoteza nulă atunci când ea este falsă (eroare de tipul II).

Probabilitatea (riscul) comiterii unei erori de tipul I este α.

Probabilitatea (riscul) comiterii unei erori de tipul II este β.

α se mai numeşte şi nivel/prag de semnificaţie; $1 - \alpha$ se mai numeşte nivelul de încredere al testului şi, exprimat în procente, reprezintă probabilitatea de garantare a rezultatelor testului; $1 - \beta$ se numeşte puterea testului statistic. Pentru a controla riscul este bine să folosim valori cât mai mici ale lui α (0,01; 0,05).

Scanned with OKEN Scanner

Regulile de decizie în cazul testelor de semnificație

Tabelul 4.6.

	H_0 adevărată	H_0 falsă
Nu se respinge H_0 (se acceptă H_0)	Decizie corectă	Eroare de tipul II Probabilitatea (eroare de tip II) = β
Se respinge H_0	Eroare de tipul I Probabilitatea (eroare de tip I) = α	Decizie corectă

Aplicarea unui test statistic de testare a ipotezelor statistice presupune întotdeauna parcurgerea următorilor paşi:

- formularea ipotezei nule şi a ipotezei alternative, care se referă la parametrul statistic al populației generale sau la legea de repartiție;
- se stabileşte testul statistic ce funcționează drept criteriu de acceptare sau respingere a ipotezei nule;
- se stabileşte regiunea critică şi pragul de semnificație;
- se formulează principalele presupuneri despre populația eşantionată;
- se calculează valoarea testului statistic, pe baza datelor culese de la unitățile statistice ale eşantionului;
- se formulează concluziile (se iau deciziile).

Vom exemplifica aplicarea testelor de verificare a ipotezelor în cazul testării ipotezei privind media populației generale (μ), în cazul eşantioanelor de volum mare ($n > 30$). Putem avea următoarele forme ale testului:

a) Testăm dacă media colectivității generale (μ) este egală cu o valoare specificată (μ_0), cu alternativa ca media să fie diferită de μ_0.

Testul bilateral:

H_0: $\mu = \mu_0$ (4.106.)

H_1: $\mu \neq \mu_0$ ($\mu < \mu_0$ sau $\mu > \mu_0$) (4.107.)

În acest caz, regiunea critică va fi stabilită în ambele capete ale distribuției (stânga-dreapta, vezi Figura 4.6.)

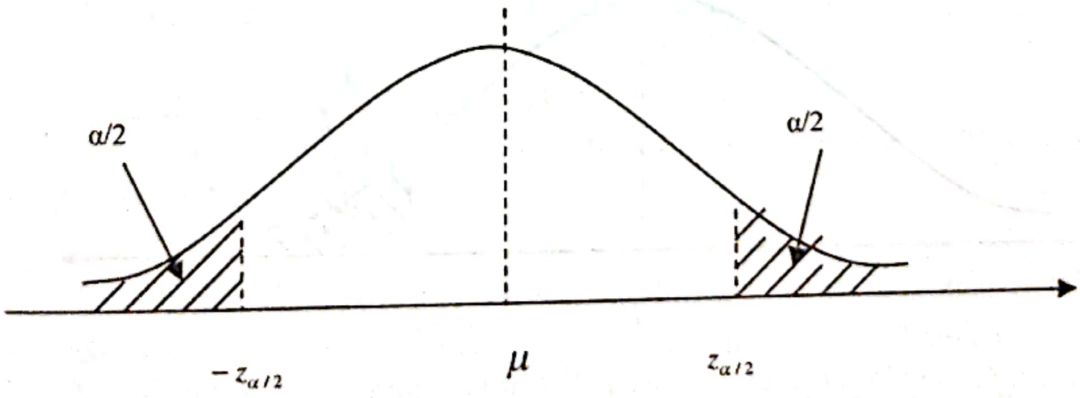


Figura 4.6. Stabilirea regiunii critice în cazul testului bilateral

Scanned with OKEN Scanner

b) Testăm dacă media colectivității generale (μ) este egală cu o valoare specificată (μ_0), cu alternativa ca media să fie mai mare decât μ_0.

Test unilateral dreapta:

H_0: $\mu=\mu_0$ (4.108.)

H_1: $\mu>\mu_0$ (4.109.)

În acest caz, regiunea critică se stabileşte numai în partea stângă a distribuţiei (Figura 4.7.).

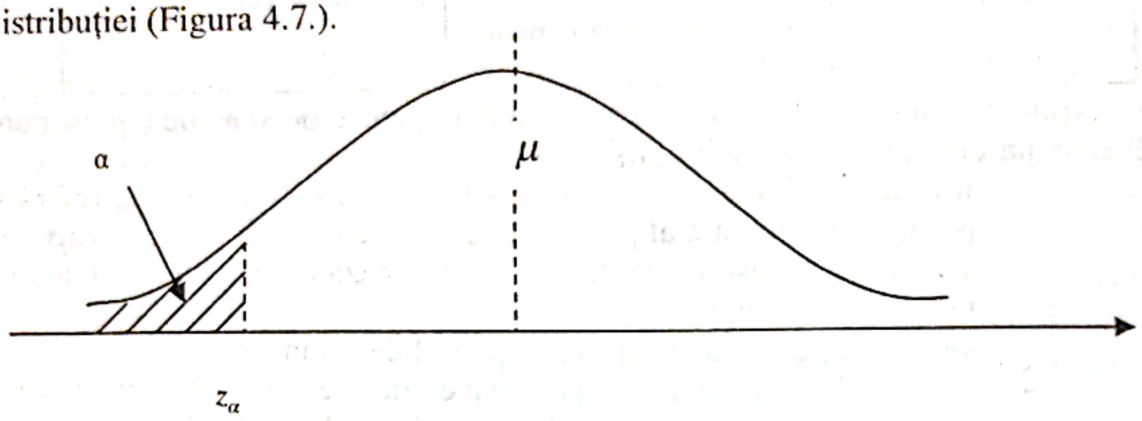


Figura 4.7. Stabilirea regiunii critice în cazul testului unilateral dreapta

c) Testăm dacă media colectivității generale (μ) este egală cu o valoare specificată (μ_0), cu alternativa ca media să fie mai mică decât μ_0.

Test unilateral stânga:

H_0: $\mu=\mu_0$ (4.110.)

H_1: $\mu<\mu_0$ (4.111.)

În acest caz, regiunea critică se stabileşte numai în partea dreaptă a distribuţiei (Figura 4.8.).

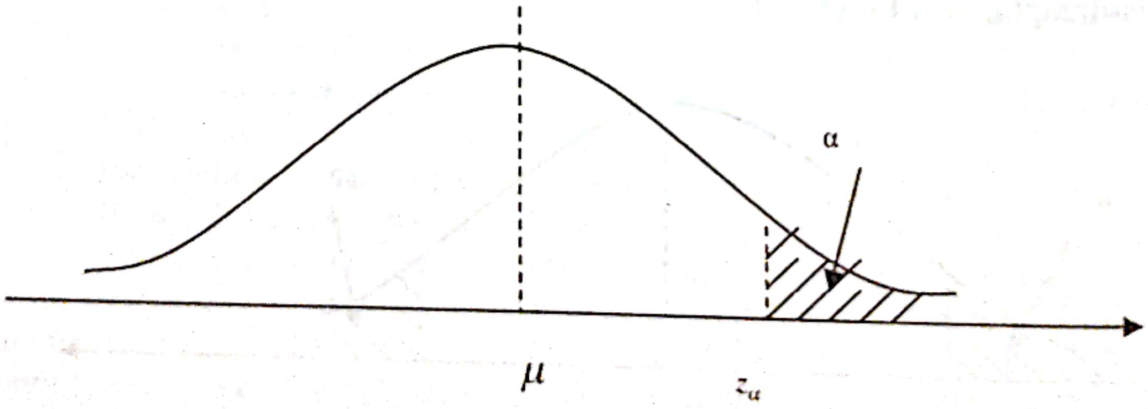


Figura 4.8. Stabilirea regiunii critice în cazul testului unilateral stânga

Scanned with OKEN Scanner

Exemplul 4.19. Într-o cercetare prin sondaj aleatoriu privitoare la transportul în comun, au fost selectate 100 de persoane pentru care s-a calculat valoarea medie a biletelor cumpărate într-o săptămână pentru transport în comun urban de 110 mii lei, cu o abatere medie pătratică de 60 mii lei.

a) Testați ipoteza nulă, aceea conform căreia valoarea medie a biletelor cumpărate într-o săptămână, în colectivitatea generală, este $\mu_0 = 100$ mii lei, cu ipoteza alternativă $\mu_0 > 100$ mii lei, utilizând un nivel de încredere $1 - \alpha = 0{,}95$ (probabilitatea $(1 - \alpha) \cdot 100 = 95\%$).

b) Testați ipoteza nulă $\mu_0 = 100$ mii lei cu ipoteza alternativă $\mu_0 \neq 100$ mii lei, cu aceeași probabilitate. Interpretați rezultatele.

Rezolvare:

a) H_0: $\mu = \mu_0 = 100$ mii lei;

Ha: $\mu > 100$ mii lei.

Testul statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{110 - 100}{60/10} = 1{,}67$

Nivelul de încredere: $1 - \alpha = 0{,}95$.

Pragul de semnificație: $\alpha = 0{,}05$

$z_\alpha = z_{0{,}05} = 1{,}645$

Cum $z_{calc} > z_\alpha$, rezultă că ipoteza nulă este respinsă ($\mu_0 = 100$ mii lei) și se acceptă ipoteza alternativă ($\mu_0 > 100$ mii lei).

b) H_0: $\mu = \mu_0 = 100$ mii lei;

Ha: $\mu \neq 100$ mii lei.

Test statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{110 - 100}{60/10} = 1{,}67$.

Nivelul de încredere $1 - \alpha = 0{,}95$.

Pragul de semnificație $\alpha = 0{,}05$, din care rezultă $\alpha/2 = 0{,}025$ (deoarece avem test bilateral).

$z_{\alpha/2} = z_{0{,}025} = 1{,}96$

Cum $z_{calc} < z_{\alpha/2}$, (1,67 < 1,96), rezultă că se acceptă ipoteza nulă (valoarea medie a biletelor cumpărate nu este semnificativ diferită de 100 mii lei).

Exemplul 4.20. Un producător de baterii dorește să verifice dacă procentul bateriilor defecte este mai mic de 5%. Presupunem că sunt selectate aleatoriu 300 de

Scanned with OKEN Scanner

baterii, fiecare dintre acestea este testată şi că sunt găsite 10 baterii defecte. Oferă aceste informaţii suficiente dovezi că procentul bateriilor defecte este mai mic de 5%? Utilizaţi o probabilitate de 99% de garantare a rezultatelor.

Rezolvare:

H_0: $\pi = 0{,}05$;

Ha: $\pi < 0{,}05$.

Testul statistic: $z = \dfrac{p - \pi}{\sigma_p} = \dfrac{w - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

$p = 10 / 300 = 0{,}03$;

$\sigma_p = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0{,}03 \cdot 0{,}97}{300}} = 0{,}0098 \cong 0{,}01$;

$z = \dfrac{0{,}03 - 0{,}05}{0{,}01} = -2$

Nivelul de încredere $1 - \alpha = 0{,}99$.

Pragul de semnificaţie $\alpha = 0{,}01$

$z_\alpha = z_{0{,}01} = 2{,}33$

Cum $|z_{calc}| < z_\alpha$, $(2 < 2{,}33)$, rezultă că ipoteza nulă nu se respinge, aşadar nu avem suficiente dovezi pentru a afirma că procentul bateriilor defecte este mai mic de 5%.

Scanned with OKEN Scanner

CAPITOLUL 5

ANALIZA STATISTICĂ A SERIILOR DE DATE MULTIVARIATE

Cele mai multe dintre metodele şi tehnicile statistice prezentate până acum au vizat descrierea sau analiza unor serii statistice univariate, comportamentul unei singure variabile statistice. Cu ajutorul metodelor statisticii descriptive am examinat modul de distribuţie al datelor sau am caracterizat valorile individuale în contextual unui set de date.

În acest capitol vom prezenta modul în care pot fi aplicate procedeele statistice pentru analiza corelată a două sau mai multe variabile statistice, fie că ele sunt corelate la un anumit moment de timp sau că se modifică în mod corelat de-a lungul unei perioade de timp. Această abordare este deosebit de utilă în studierea celor mai multe dintre fenomenele şi procesele economico-sociale: în afaceri, în marketing, în domeniul comportamental etc., rezultatul activităţii este influenţat de o paletă largă de factori interrelaţionali. Cu ajutorul metodelor şi tehnicilor statistice pot fi identificate asemenea dependenţe între multiplele fenomene, pot fi descrise cantitativ legăturile, se poate măsura intensitatea cu care se produc, se poate previziona nivelul unei variabile dependente, pentru anumite valori, fixate, ale variabilelor independente. Aceste aspecte sunt deosebit de utile în procesul de programare, de planificare a producţiei, în controlul stocurilor, în dezvoltarea strategiilor de marketing.

5.1. Conceptul de legătură statistică. Clasificarea legăturilor statistice

Dacă două variabile se modifică în acelaşi sens sau în sensuri contrare, am putea trage concluzia că între ele există o interdependenţă, o legătură, că ar putea fi corelate. Totuşi, s-ar putea ca modificarea unei variabile să nu implice neapărat o modificare în nivelul celeilalte variabile, ci ambele variabile să se modifice în funcţie de o a treia variabilă[1]. De exemplu, se constată că într-o staţiune de pe litoral, în luna iulie s-a înregistrat o creştere a vânzărilor de creme de protecţie solară, dar şi a numărului de turişti sosiţi în staţiune. La prima vedere, s-ar părea că există o legătură între cele două variabile, însă în realitate aceste caracteristici nu ar

[1] Emilia Ţiţan – *Statistică. Teorie şi aplicaţii în sectorul terţiar*, Editura Meteor Press, Bucureşti, 2003.

Scanned with OKEN Scanner

fi înregistrat sporuri ale nivelului lor, dacă nu s-ar fi atins temperaturi ridicate şi o vreme însorită. Aşadar, ambele variabile se modifică de fapt în pas cu o a treia variabilă: vremea din luna iulie.

Înainte de a analiza o legătură între două variabile statistice, trebuie deci să stabilim dacă această legătură există în mod real, dacă nu este cumva o falsă relaţie.

Între fenomenele din natură şi societate există mai multe tipuri de legături. Ele se împart în două grupe mari: legături deterministe şi legături nedeterministe sau stochastice.

I. ***Legăturile deterministe*** sunt legături univoc determinate, cu un caracter mai strict, mai rigid, ce apar, de regulă, între fenomenele din fizică, chimie etc. (ştiinţele naturii). Caracteristic acestui tip de legătură este faptul că producerea fenomenului cauză induce în mod necesar şi producerea fenomenului efect, iar unei valori a caracteristicii cauzale îi corespunde o singură valoare, strict determinată, a caracteristicii rezultative. Legăturile deterministe pot fi identificate şi analizate pe cale experimentală, în laborator, validitatea lor dovedindu-se în fiecare caz individual în parte.

II. ***Legăturile nedeterministe (stochastice sau statistice)*** sunt, în general, caracteristice fenomenelor economico-sociale. Aceste fenomene sunt rezultatul acţiunii unui număr mare de factori de influenţă. Între aceşti factori regăsim şi unii de natură sistematică, care pot fi identificaţi, dar şi factori de natură aleatorie, care, de multe ori, nu pot fi stabiliţi şi cuantificaţi. De aceea, specific legăturilor nedeterministe este faptul că unei valori a caracteristicii cauzale îi corespunde o distribuţie de valori a caracteristicii rezultative, ce pot fi tratate ca variabile aleatorii şi studiate cu metode statistice. Legăturile statistice pot fi identificate doar prin studierea unui număr mare de unităţi statistice de la care s-au înregistrat variabilele urmărite şi nu se verifică în mod cert în fiecare caz individual în parte.

La rândul lor, legăturile statistice se pot clasifica după următoarele criterii:

a) după *natura variabilelor* implicate, putem avea:
   i. legături între variabile numerice (numite ***corelaţii***);
   ii. legături între variabile calitative sau între variabile numerice şi nenumerice (numite ***asocieri***).

De exemplu, între suprafaţa comercială a unor supermarketuri şi profitul din vânzări obţinut de ele există o corelaţie. Între calificativul primit de nişte studenţi la matematică şi nota obţinută la examenul de statistică există o asociere.

b) după *numărul variabilelor* între care se stabileşte legătura, avem:
   i. ***legături simple*** (între două variabile, dintre care una este cauză şi alta efect);
   ii. ***legături multiple*** (între trei sau mai multe variabile, din care una este efect şi cel puţin două sunt variabile cauzale).

Scanned with OKEN Scanner

De exemplu, dacă studiem legătura între vechimea în muncă a unor salariaţi şi salariul lor lunar, atunci avem o legătură simplă. Dacă însă analizăm relaţia dintre profesia, vechimea în muncă şi salariul lunar al angajaţilor unei companii, avem o legătură multiplă.

c) după *direcţia* sau *sensul legăturii*, avem:
   i. ***legături directe*** (dacă modificarea variabilei cauzale şi a celei rezultative se produce în acelaşi sens: ambele cresc sau ambele scad);
   ii. ***legături inverse*** (când variabilele cauză şi efect se modifică în sensuri diferite: când prima înregistrează o creştere, a doua înregistrează o scădere sau invers).

De exemplu, între nivelul productivităţii muncii şi nivelul salariului este o legătură directă; între numărul învoirilor unor angajaţi şi producţia realizată există o legătură inversă.

d) după *forma legăturii*, avem:
   i. ***legături liniare***: atunci când modificarea variabilei cauzale determină modificarea uniformă a variabilei efect, acţiunea celorlalţi factori de influenţă fiind considerată constantă; în acest caz, tendinţa legăturii se poate modela cu ajutorul unei funcţii analitice liniare;
   ii. ***legături neliniare*** (***curbilinii***): atunci când tendinţa legăturii dintre variabilele studiate se poate modela cu ajutorul unei funcţii de tip neliniar (parabola de gradul doi, exponenţială, polinomială de ordin superior etc.).

e) după *modul de manifestare în plan temporal*, putem avea:
   i. ***legături sincrone*** (***concomitente***), atunci când modificarea variabilei rezultative are loc aproape în acelaşi timp cu modificarea variabilei cauzale (de exemplu: legătura dintre vechimea în muncă şi nivelul salariului);
   ii. ***legături asincrone*** (sau cu ***decalaj***), atunci când modificarea variabilei rezultative are loc după o perioadă de timp de la schimbarea variabilei cauzale (de exemplu: legătura dintre investiţiile în maşini, utilaje noi, moderne, pe de o parte şi creşterea productivităţii muncii, pe de altă parte).

Metodele de analiză a legăturilor dintre variabilele statistice se împart în două mari categorii:

a) ***metode elementare*** (simple) de analiză a legăturilor, care permit stabilirea existenţei legăturii, a direcţiei şi a formei acesteia;
b) ***metode complexe***, din care fac parte **metodele parametrice** şi **metodele neparametrice**; dintre metodele parametrice amintim *metoda regresiei*, prin care se modelează tendinţa legăturii cu ajutorul unei funcţii analitice, pe baza căreia se poate previziona nivelul variabilei rezultative în funcţie

Scanned with OKEN Scanner

de valorile fixate ale variabilei cauzale şi *metoda corelaţiei*, ce permite măsurarea intensităţii legăturii şi aprecierea sensului acesteia.

5.2. Metode elementare de analiză a legăturilor dintre variabilele statistice

Metodele elementare de analiză a legăturilor dintre fenomenele economico-sociale sunt: metoda grafică, metoda seriilor paralele (sau interdependente), metoda grupării şi metoda tabelului de corelaţie.

a) Metoda grafică

Reprezentarea grafică a distribuţiilor de frecvenţe bidimensionale (a seriilor bidimensionale) se numeşte corelogramă sau diagrama norilor de puncte. Ea a fost prezentată în capitolul 2.

Corelograma permite, prin interpretarea modului de amplasare a punctelor în spaţiul determinat de cele două axe, a existenţei, direcţiei (sensului) şi a formei legăturii. Este cea mai des utilizată metodă elementară de analiză a legăturilor dintre variabile. Pe baza ei putem efectua următoarele comentarii:

Existenţa legăturii:

- dacă punctele, care reprezintă unităţile statistice de la care s-au înregistrat cele două variabile urmărite, sunt aşezate într-o zonă cu o anumită formă, bine conturată, din spaţiul graficului, atunci putem concluziona că există legătură între variabile. Dacă punctele sunt, dimpotrivă, risipite în întreg spaţiul definit de cele două axe, atunci înseamnă că nu există legătură sau este de intensitate foarte mică, practic nesemnificativă [Figura 5.1. a), b) şi c)].

Scanned with OKEN Scanner

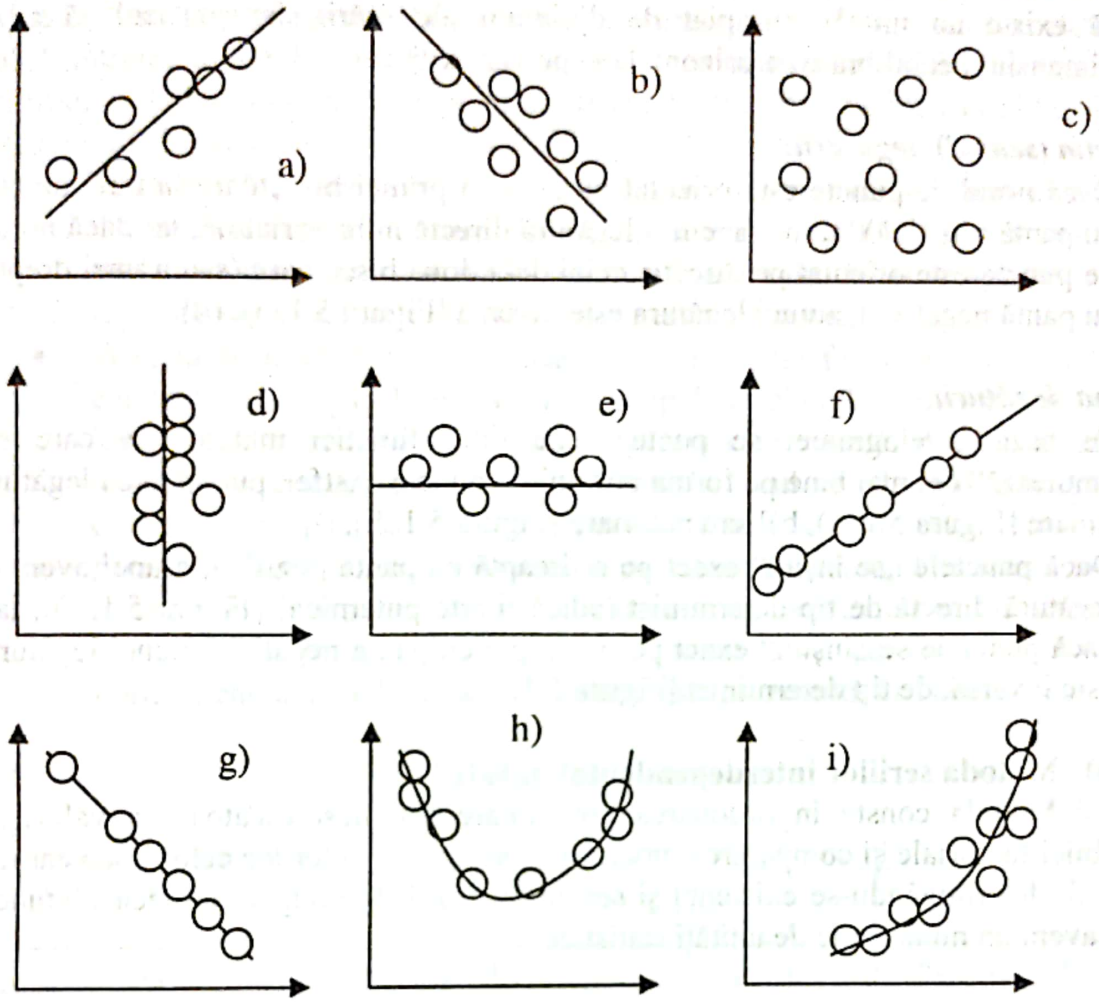


Figura 5.1. Corelograma. Tipuri de legături. a) legătură liniară directă; b) legătură liniară inversă; c), d), e) absența legăturii; f) legătură liniară deterministă directă; g) legătură liniară deterministă inversă; h), i) legături neliniare

- Dacă zona în care se concentrează punctele de coordonate (x_i, y_i) (unde x_i este valoarea caracteristicii cauzale la unitatea statistică i, iar y_i este valoarea caracteristicii rezultative la unitatea statistică i) are direcția unei drepte orizontale sau verticale, atunci nu există legătură între variabile [Figura 5.1. d), e)]. Alegerea scării de reprezentare trebuie făcută cu atenție, deoarece prin schimbarea sa pe una din axe, un nor de puncte care, inițial, era dispersat uniform în tot planul graficului, poate apoi apărea ca fiind de o formă mai bine conturată, orientată de-a lungul unei drepte paralele cu una din axe. La construirea corelogramei este permis ca originea axelor să nu fie în punctul zero, ci în valori mai mici decât valorile minime ale celor două variabile implicate în

Scanned with OKEN Scanner

analiză, alese convenabil. Este, de asemenea, recomandat ca pe cele două axe să existe un număr apropiat de diviziuni ale scării, iar graficul să aibă dimensiuni echilibrate pe orizontală şi pe verticală.

Direcţia (sensul) legăturii:

- Dacă norul de puncte este orientat pe direcţia primei bisectoare (a unei drepte cu pantă pozitivă), atunci avem o legătură directă între variabile, iar dacă norul de puncte este orientat pe direcţia celei de-a doua bisectoare (sau a unei drepte cu pantă negativă), atunci legătura este inversă [Figura 5.1. a), b)].

Forma legăturii:

- Pe baza corelogramei se poate alege tipul funcţiei matematice care se „mulează" cel mai bine pe forma norului de puncte. Astfel, putem avea legături liniare [Figura 5.1. a), b)] sau neliniare [Figura 5.1. h), i)].
- Dacă punctele „se înşiră" exact pe o dreaptă cu pantă pozitivă, atunci avem o legătură directă de tip determinist (adică foarte puternică) [Figura 5.1. f)], iar dacă punctele se „înşiră" exact pe o dreaptă cu pantă negativă, atunci legătura este inversă, de tip determinist [Figura 5.1. g)].

b) Metoda seriilor interdependente

Metoda constă în ordonarea crescătoare sau descrescătoare a valorilor variabilei factoriale şi compararea modului de variaţie a valorilor celor două caracteristici, determinându-se existenţa şi sensul legăturii. Metoda se utilizează atunci când avem un număr mic de unităţi statistice.

c) Metoda grupării

Metoda presupune gruparea unităţilor statistice pe grupe după valorile variabilei cauzale, în timp ce, pentru variabila rezultativă, se calculează indicatori derivaţi (medii, relativi). De exemplu, se pot determina medii condiţionate ale variabilei efect, pe grupe după variabila cauzală. În urma comparării variaţiei celor două variabile, se identifică existenţa şi sensul legăturii. Aplicarea acestei metode se poate realiza atunci când avem un număr mare de unităţi statistice, dar pentru a identifica corect legătura trebuie să avem un număr suficient de mare de grupe.

d) Metoda tabelului de corelaţie

Este o metodă aplicată în cazul unor grupări combinate a datelor statistice, după două variabile numerice. La elaborarea tabelului de corelaţie se ţine seama de principiile şi regulile de construire a graficului de corelaţie. Astfel, variantele numerice sau intervalele de variaţie ale variabilei cauzale se trec în capetele coloanelor, în ordine crescătoare, de la stânga la dreapta (pe orizontală), în timp ce

Scanned with OKEN Scanner

variantele numerice sau intervalele de variație ale variabilei rezultative se trec în capetele rândurilor (pe verticală), în ordine descrescătoare. În rubricile tabelului cu dublă intrare, aflate la întretăierea liniilor cu coloanele, se trec frecvențele de apariție. După modul în care se amplasează frecvențele în interiorul tabelului, obținem informații privind:

- ***existența legăturii***: dacă frecvențele sunt concentrate în jurul uneia din diagonalele tabelului, atunci înseamnă că există legătură; dacă frecvențele sunt dispersate, împrăștiate în interiorul întregului tabel, atunci nu există legătură sau este foarte slabă (practic nesemnificativă);
- ***direcția (sensul) legăturii***: dacă cele mai multe frecvențe de apariție se concentrează în jurul diagonalei principale a tabelului (între colțul stânga jos şi cel din dreapta sus) atunci legătura este directă; dacă, dimpotrivă, majoritatea frecvențelor sunt amplasate în jurul celeilalte diagonale (din colțul stânga sus spre cel din dreapta jos), atunci legătura este inversă;
- ***forma legăturii***: dacă frecvențele se distribuie de-a lungul unei linii drepte, atunci există o legătură liniară; în caz contrar, frecvențele pot fi amplasate într-o zonă de formă neliniară (legătură neliniară sau curbilinie).

Forma generală a tabelului de corelație este prezentată în Tabelul 5.1.

Tabelul de corelație

Tabelul 5.1.

Valori sau intervale pt. variabila efect Y	Valori sau intervale pt. variabila cauzală X						Total unități
	x_1	x_2	...	x_i	...	x_r	
0	1	2	3	4	5	6	7
y_p	n_{p1}	n_{p2}	...	n_{pi}	...	n_{pr}	$n_{p.}$
...	...	...	...	...	...	...	...
y_j	n_{j1}	n_{j2}	...	n_{ji}	...	n_{jr}	$n_{j.}$
...							
y_2	n_{21}	n_{22}	...	n_{2i}	...	n_{2r}	$n_{2.}$
y_1	n_{11}	n_{12}	...	N_1	...	n_{1r}	$n_{1.}$
Total unități	$n_{.1}$	$n_{.2}$	...	$n_{.i}$	...	$n_{.r}$	$n_{..}$

unde:

$$n_{i.} = \sum_{j=1}^{p} n_{ij} \qquad (5.1.)$$

$$n_{.j} = \sum_{i=1}^{r} n_{ij} \qquad (5.2.)$$

Scanned with OKEN Scanner

$$n.. = n = \sum_{i=1}^{r}\sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{p} n_{.j} \qquad (5.3.)$$

Dacă variabilele între care se stabileşte legătura sunt calitative, atunci tabelul care redă rezultatele sistematizării unităţilor statistice după aceste variabile se numeşte tabel de asociere sau de contingenţă şi va fi prezentat mai pe larg în secţiunea 5.5.1. a prezentului capitol.

5.3. Metode parametrice de analiză a legăturilor dintre variabilele statistice

Metoda regresiei este utilizată pentru a previziona nivelul unei variabile rezultative Y în funcţie de valorile uneia sau mai multor variabile cauzale, folosind un model matematic numit funcţie de regresie. Această metodă permite exercitarea unui control asupra outputului, pentru anumite valori, strict determinate, ale inputului.

Aplicarea metodei regresiei presupune stabilirea în prealabil a *relaţiei de cauzalitate* dintre variabilele studiate, adică identificarea cauzei şi a efectului (a sensului implicaţiei). Sunt cazuri în care stabilirea variabilei cauzale şi a variabilei rezultative nu este chiar aşa de facilă, întrucât implicaţia dintre ele ar putea avea un dublu sens (implicaţie reciprocă, interdependenţă). În acest caz se poate alege unul dintre sensurile implicaţiei, acela care este mai evident sau acela care interesează pe cercetător.

Relaţia de dependenţă dintre variabila sau variabilele cauzale, pe de o parte, şi cea rezultativă, pe de altă parte, se poate scrie sub forma unei ecuaţii, de tipul:

$$y = f(x) = f(x_1, x_2, ..., x_n) \qquad (5.4.)$$

unde

$x = (x_1, x_2, ..., x_n)$ este variabila cauzală, factorială, independentă, exogenă sau explicativă, iar

Y este variabila efect, rezultativă, dependentă, endogenă sau explicată.

Relaţia (5.4.) reprezintă modelul teoretic (deterministic) ce exprimă dependenţa dintre variabile. Între toţi factorii de influenţă $x = (x_1, x_2, ..., x_n)$ există, însă, şi unii de natură aleatorie, a căror acţiune nu poate fi întotdeauna identificată şi cuantificată numeric. Din această cauză, prin înlocuirea în modelul de regresie a valorilor variabilei/variabilelor cauzale cu unele strict determinate, nu vom regăsi exact valorile reale ale variabilei rezultative. Modelul teoretic se

Scanned with OKEN Scanner

înlocuieşte cu un model probabilistic [relaţia (5.5.)], care ţine seama şi de posibilităţile apariţiei unor factori aleatorii.

$$y = f(x) + \varepsilon = f(x_1, x_2, ..., x_n) + \varepsilon \qquad (5.5.)$$

unde ε este *eroare aleatorie*, sau *variabilă reziduu* şi reflectă influenţa factorilor aleatorii.

Dacă luăm în studiu o singură variabilă factorială, atunci avem *regresie simplă*, iar dacă ţinem cont de doi sau mai mulţi factori de influenţă, atunci avem *regresie multiplă*.

5.3.1. Metoda regresiei liniare simple

În încercarea de a previziona cît mai exact nivelul variabilei rezultative pe baza valorilor variabilei cauzale, trebuie aleasă, pe baza graficului de corelaţie, funcţia de regresie cea mai potrivită, care se „mulează" cel mai bine pe forma norului de puncte ce redă tendinţa legăturii.

Dacă funcţia de regresie aleasă este una liniară, atunci regresia se numeşte ***regresie liniară***, iar dacă se recurge la o funcţie curbilinie (neliniară), regresia se va numi ***regresie neliniară***.

Modelul probabilistic general pentru regresia liniară simplă este:

$$y_i = a + b \cdot x_i + \varepsilon_i \qquad (5.6.)$$

unde

(x_i, y_i) reprezintă valorile numerice ale variabilelor cauză şi efect înregistrate la nivelul unităţii statistice *i*;

a = punctul de intersecţie al dreptei de regresie cu axa Oy;

b = panta dreptei, se mai numeşte şi *coeficient de regresie* şi arată cu câte unităţi de măsură se modifică *Y* dacă *X* se modifică cu o unitate de măsură;

ε_i = componenta reziduală (eroare aleatorie) pentru unitatea statistică *i*.

Aşadar, valoarea reală y_i a caracteristicii *Y* cuprinde o *componentă teoretică, previzionabilă* ($\hat{y}_i$), care se poate determina pe baza modelului pentru o anumită valoare x_i, şi o *componentă aleatorie (reziduală)*, numită şi *eroarea aleatorie*, (ε_i), reprezentând acea parte din valoarea reală a lui *Y* care nu se poate cuantifica:

$$y_i = \hat{y}_i + \varepsilon_i \qquad (5.7.)$$

$$\hat{y}_i = a + b \cdot x_i, \; i = \overline{1,n} \qquad (5.8.)$$

Scanned with OKEN Scanner

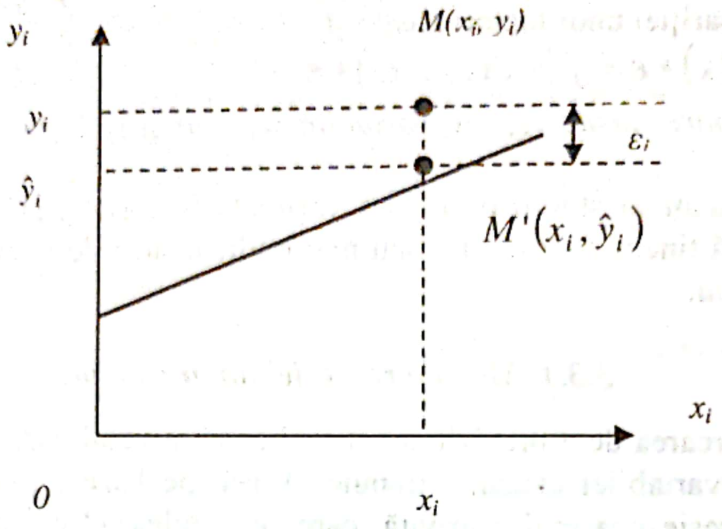


Figura 5.2. Dreapta de regresie

Pentru a asigura obținerea unor estimatori de calitate ridicată ai parametrilor modelului de regresie liniară simplă, trebuie emise și respectate următoarele ***ipoteze***, la nivelul colectivității generale:

1. valorile variabilei factoriale X, fie că sunt fixate sau că sunt variabile aleatorii, sunt independente statistic de variabilele aleatorii, reziduale, ε_i;
2. datele privind variabilele Y și X sunt considerate corecte, neafectate de erori de observare sau măsurare;
3. valorile variabilei reziduale ε_i urmează o distribuție normală, de medie zero ($\mu(\varepsilon_i)=0, i=\overline{1,n}$) și dispersie constantă și nenulă, pentru toate valorile X_i. ($\sigma_\varepsilon^2 = const, \sigma_\varepsilon^2 \neq 0$). Cu alte cuvinte, întrucât distribuția variabilei reziduu este independentă de valorile variabilelor explicative, nici dispersia perturbației σ_ε^2 nu diferă semnificativ în raport cu valorile X_i, ceea ce indică o stabilitate relativă a legăturii dintre variabila rezultativă și variabilele factoriale (omoscedasticitate);
4. variabilele aleatorii ε_i sunt statistic independente una de alta, adică $\rho_{\varepsilon i \varepsilon j} = 0$, $i \neq j$. Spunem deci, în acest caz, că valorile variabilei reziduale nu sunt autocorelate (nu există corelație între reziduuri). Această ipoteză indică și corectitudinea realizării observațiilor, a independenței acestora.

Scanned with OKEN Scanner

5.3.1.1. Determinarea parametrilor funcției de regresie liniară

Pentru determinarea dreptei celei mai potrivite pentru modelarea tendinței legăturii dintre cele două variabile (adică a dreptei care trece cel mai aproape de cele mai multe din punctele (x_i, y_i)) se impune calcularea parametrilor a și b ai dreptei de regresie. Acest lucru se realizează prin mai multe metode, dar cea mai cunoscută este ***metoda celor mai mici pătrate***.

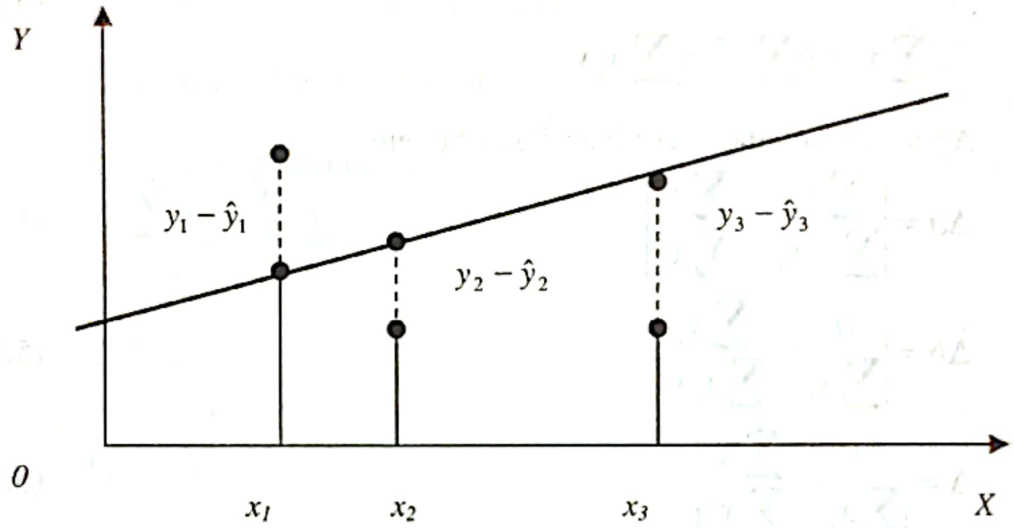


Figura 5.3. Metoda celor mai mici pătrate

Dreapta de regresie determinată prin metoda celor mai mici pătrate este amplasată pe graficul de corelație astfel încât suma pătratelor distanțelor verticale între punctele $M(x_i, y_i)$ și dreapta de regresie este minimă. Altfel spus, dreapta de regresie care trece cel mai aproape de punctele $M(x_i, y_i)$ este cea pentru care suma pătratelor diferențelor între valorile reale și cele teoretice ale lui Y este minimă.

$$\sum d_i^2 = \sum (y_i - \hat{y}_i)^2 \rightarrow \text{minim} \qquad (5.9.)$$

Motivul pentru care se iau în considerare pătratele distanțelor d_i și nu distanțele ca atare este că unele valori ajustate sunt mai mari sau mai mici decât valorile reale, deci și distanțele sunt pozitive sau negative și, prin însumarea lor, aceste distanțe s-ar putea compensa reciproc. Într-adevăr, unele puncte $M(x_i, y_i)$ sune plasate deasupra dreptei de regresie, altele – dedesubtul ei.

Scanned with OKEN Scanner

Din relațiile (5.8.) și (5.9.) rezultă că:

$$\sum d_i^2 = \sum \left(y_i - a - b \cdot x_i\right)^2 = \sum \varepsilon_i^2 \rightarrow \text{minim} \qquad (5.10.)$$

Pentru determinarea parametrilor a și b trebuie derivată expresia din relația (5.10.) în raport cu a și b și apoi derivatele – egalate cu zero. Rezultă următorul sistem de două ecuații simultane, cu necunoscutele a și b:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \qquad (5.11.)$$

Aplicând metoda determinanților, obținem:

$$\Delta a = \begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix} \qquad (5.12.)$$

$$\Delta b = \begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} \qquad (5.13.)$$

$$\Delta = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} \qquad (5.14.)$$

$$a = \frac{\Delta a}{\Delta} = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad (5.15.)$$

$$b = \frac{\Delta b}{\Delta} = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad (5.16.)$$

a poate lua valori atât pozitive, cât și negative. b, fiind panta dreptei, indică direcția legăturii, astfel:

- Dacă $b > 0$, atunci legătura este directă;
- Dacă $b < 0$, atunci legătura este inversă;
- Dacă $b = 0$, atunci nu există legătură liniară, variabilele X și Y sunt independente (Figura 5.4.).

Scanned with OKEN Scanner

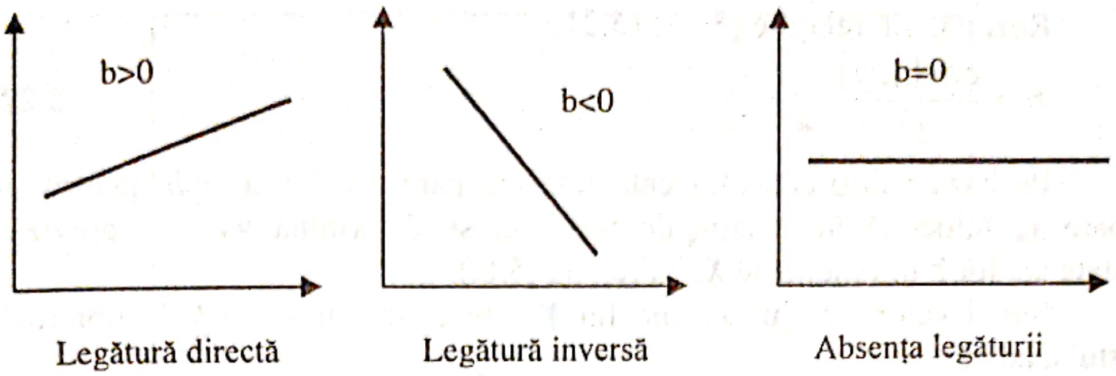


Figura 5.4. Corelația dintre semnul parametrului *b* și direcția legăturii

Dacă împărțim prima ecuație din sistemul (5.11.) la *n*, obținem:

$$a + b \cdot \frac{\sum x_i}{n} = \frac{\sum y_i}{n} \tag{5.17.}$$

sau:

$$a + b \cdot \bar{x} = \bar{y} \tag{5.18.}$$

Relația (5.18.) arată că dreapta de regresie trece printr-un punct fix, de coordonate $(\bar{x}, \bar{y})$. Prelucrând expresia (5.16.), obținem:

$$b = \frac{\Delta b}{\Delta} = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} = \frac{\frac{\sum x_i y_i}{n} - \frac{\sum x_i}{n} \cdot \frac{\sum y_i}{n}}{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} =$$

$$= \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2} \tag{5.19.}$$

Numărătorul expresiei (5.19.) reprezintă un indicator numit *covarianță* și care va fi detaliat în secțiunile următoare ale prezentului capitol.

$$\mathrm{cov}(x, y) = \overline{xy} - \bar{x} \cdot \bar{y} = \frac{\sum x_i y_i}{n} - \frac{\sum x_i}{n} \cdot \frac{\sum y_i}{n} \tag{5.20.}$$

Numitorul expresiei (5.19.) este dispersia variabilei *X*:

$$s_x^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 \tag{5.21}$$

Scanned with OKEN Scanner

Rezultă, din relațiile (5.19)-(5.21):

$$b = \frac{\text{cov}(x, y)}{s_x^2} \qquad (5.22.)$$

Pe baza valorilor astfel calculate ale parametrilor a și b, pentru fiecare valoare x_i, înlocuită în ecuația de regresie, se determină valorile previzionate, ajustate ale lui Y în funcție de X, cu relația (5.8.).

Suma valorilor ajustate ale lui Y este egală cu suma valorilor reale ale acestuia, adică:

$$\sum_{i=1}^{n} \hat{y}_i \approx \sum_{i=1}^{n} y_i \qquad (5.23.)$$

Exemplul 5.1. Din totalul fabricilor de încălțăminte, s-au ales aleatoriu 10 dintre acestea, de la care s-au înregistrat numărul de comenzi și profitul realizat (Tabelul 5.2.):

Numărul de comenzi și profitul realizat pentru fabricile de încălțăminte

Tabelul 5.2.

Fabrica de încălțăminte	Număr de comenzi (x_i)	Profit (mld. lei) (y_i)
0	1	2
A	31	330
B	43	400
C	15	152
D	17	230
E	20	225
F	14	132
G	18	280
H	40	500
I	27	300
J	38	420
Total	**263**	**2969**

Corelograma pentru datele din Tabelul 5.2. se prezintă în Figura 5.5.

Scanned with OKEN Scanner

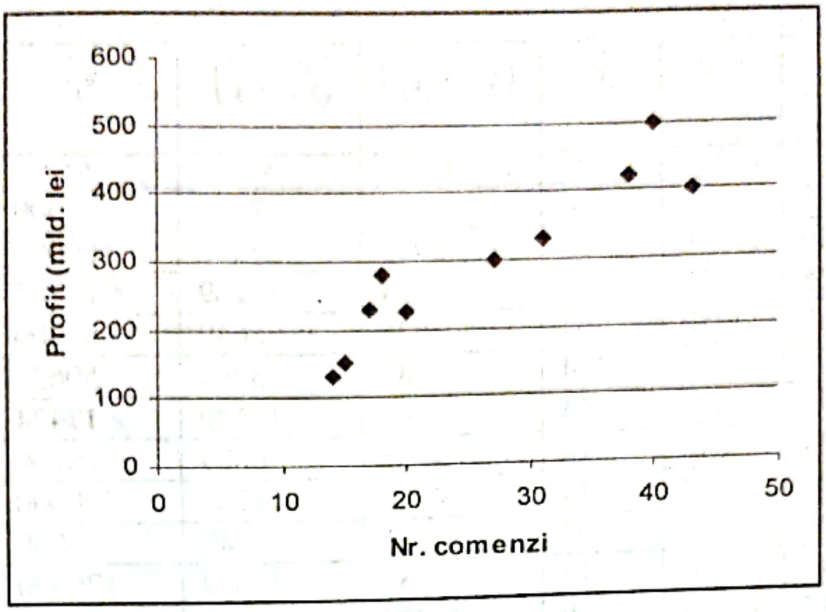


Figura 5.5. Dependența profitului de numărul de comenzi

Numărul de comenzi reprezintă cauza (X), iar profitul – efectul (Y). În urma elaborării corelogramei (Figura 5.5.), se observă că între variabile există o legătură directă liniară. Vom aplica, atunci, modelul de regresie liniar pentru $n = 10$, iar prin aplicarea metodei celor mai mici pătrate, rezultă următoarele valori pentru a şi b:

$$a = \frac{\Delta a}{\Delta} = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{2969 \cdot 8017 - 263 \cdot 89068}{80170 - 263^2} = 34{,}32$$

$$b = \frac{\Delta b}{\Delta} = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{890680 - 263 \cdot 2969}{80170 - 263^2} = \frac{109833}{11001} = 9{,}98$$

Ecuația de regresie va fi:

$$\hat{y}_i = 34{,}32 + 9{,}98 \cdot x_i$$

Cum $b > 0$, înseamnă că legătura este directă; dacă numărul comenzilor creşte cu o comandă, atunci profitul va creşte cu 9,98 mld. lei. Calculele intermediare se găsesc în Tabelul 5.3.

Scanned with OKEN Scanner

Tabelul 5.3.

Fabrica	x_i^2	$x_i y_i$	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$(\hat{y}_i - \bar{y})^2$	y_i^2	$(y_i - \bar{y})^2$
0	1	2	3	4	5	6	7
A	961	10230	344	196	2209	108900	1089
B	1849	17200	463	3969	27556	160000	10609
C	225	2280	184	1024	12769	23104	21025
D	289	3910	204	676	8649	52900	4489
E	400	4500	234	81	3969	50625	5184
F	196	1848	174	1764	15129	17424	27225
G	324	5040	214	4356	6889	78400	289
H	1600	20000	434	4356	18769	250000	41209
I	729	8100	304	16	49	90000	9
J	1444	15960	414	36	13689	176400	15129
Total	**8017**	**89068**	**2969**	**16474**	**109677**	**1007753**	**126.257**

Dacă datele au fost sistematizate pe grupe după X şi Y şi deci perechile de valori (x_i, y_i) se întâlnesc de n_i ori, sistemul de ecuaţii simultane devine:

$$\begin{cases} a\sum_{i=1}^{r} n_i + b\sum_{i=1}^{r} x_i n_i = \sum_{i=1}^{r} y_i n_i \\ a\sum_{i=1}^{r} x_i n_i + b\sum_{i=1}^{r} x_i^2 n_i = \sum_{i=1}^{r} x_i y_i n_i \end{cases} \qquad (5.24.)$$

Parametrii a şi b vor fi:

$$a = \frac{\Delta a}{\Delta} = \frac{\sum y_i n_i \cdot \sum x_i^2 n_i - \sum x_i n_i \cdot \sum x_i y_i n_i}{\sum n_i \sum x_i^2 n_i - \left(\sum x_i n_i\right)^2} \qquad (5.25.)$$

$$b = \frac{\Delta b}{\Delta} = \frac{\sum n_i \cdot \sum x_i y_i n_i - \sum x_i n_i \cdot \sum y_i n_i}{n\sum x_i^2 n_i - \left(\sum x_i n_i\right)^2} \qquad (5.26.)$$

Dacă datele au fost sistematizate într-un tabel cu dublă intrare şi perechile de valori (x_i, y_j) se întâlnesc de n_{ij} ori, sistemul devine:

$$\begin{cases} a\sum_{i=1}^{r}\sum_{j=1}^{p} n_{ij} + b\sum_{i=1}^{r} x_i n_{i.} = \sum_{j=1}^{p} y_j n_{.j} \\ a\sum_{i=1}^{r} x_i n_{i.} + b\sum_{i=1}^{r} x_i^2 n_{i.} = \sum_{i=1}^{r}\sum_{j=1}^{p} x_i y_j n_{ij} \end{cases} \qquad (5.27.)$$

Scanned with OKEN Scanner

Parametrii a şi b vor fi:

$$a = \frac{\sum_j y_j n_{.j} \cdot \sum_i x_i^2 n_{i.} - \sum_i x_i n_{i.} \cdot \sum_i \sum_j x_i y_j n_{ij}}{\sum_i \sum_j n_{ij} \sum_i x_i^2 n_{i.} - \left(\sum_i x_i n_{i.} \right)^2} \qquad (5.28.)$$

$$b = \frac{\sum_i \sum_j n_{ij} \sum_i \sum_j x_i y_j n_{ij} - \sum_i x_i n_{i.} \cdot \sum_j y_j n_{.j}}{\sum_i \sum_j n_{ij} \sum_i x_i^2 n_{i.} - \left(\sum_i x_i n_{i.} \right)^2} \qquad (5.29.)$$

Pentru analiza calităţii ajustării prin regresia liniară simplă, se determină abaterea medie pătratică a erorilor în eşantion:

$$s_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}} \qquad (5.30.)$$

unde:

s_ε^2 este un estimator nedeplasat al dispersiei reziduurilor $\sigma_{\varepsilon 0}^2$.

5.3.1.2. Testarea semnificaţiei parametrilor de regresie şi a validităţii modelului de regresie liniară simplă

Presupunem că datele înregistrate pentru variabilele X şi Y se referă la unităţile unui eşantion de volum n ($i = \overline{1,n}$). Avem deci n perechi de valori (x_i, y_i). Modelul de regresie liniar la nivelul eşantionului va fi:

$$y_i = a + b \cdot x_i + \varepsilon_i \qquad (5.31.)$$

cu componenta deterministă:

$$\hat{y}_i = a + b \cdot x_i, \; i = \overline{1,n} \qquad (5.32.)$$

a şi b sunt estimatorii parametrilor a_0 şi b_0 din modelul de regresie corespunzător colectivităţii generale:

$$Y_i = a_0 + b_0 \cdot X_i + \varepsilon_{0i} \qquad (5.33.)$$

Este de dorit ca a şi b să estimeze cât mai exact coeficienţii a_0 şi b_0 din modelul de regresie pentru întreaga colectivitate.

Notăm, de asemenea, cu $\bar{x}$ media variabilei X la nivelul eşantionului şi cu μ media variabilei X la nivelul colectivităţii generale. Coeficienţii sunt variabile aleatorii, cu distribuţii de eşantionare.

Scanned with OKEN Scanner

Aplicarea testelor statistice este necesară pentru a determina:

- condițiile în care estimațiile obținute la nivelul eșantionului sunt valabile și la nivelul întregii colectivități;
- dacă valoarea estimatorului este semnificativă statistic și nu e rezultatul întâmplării (al unor conjuncturi aleatoare);
- intervalul de încredere în interiorul căruia poate lua valori estimatorul obținut și probabilitatea cu care se garantează că valorile reale ale parametrilor (la nivelul colectivității totale) se regăsesc în intervalul de încredere determinat.

Am văzut că una din ipotezele modelului de regresie era aceea că variabila reziduală ε_i urmează o repartiție normală și că estimatorii a și b din modelul liniar de regresie simplă nu sunt combinații liniare ale variabilei ε_i. De aici rezultă că parametrii a și b (estimatorii) au și ei distribuții de eșantionare, ce prezintă următoarele proprietăți:

- a și b sunt estimatori nedeplasați ai parametrilor a_0 și b_0, adică $\mu(a) = a_0$ și $\mu(b) = b_0$ (media estimațiilor este egală cu parametrul populației generale);
- distribuțiile de eșantionare ale parametrilor a și b sunt normal distribuite, cu mediile a_0 și b_0 și dispersiile s_a^2 și s_b^2.

$$s_a^2 = s_\varepsilon^2 \cdot \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{5.34.}$$

$$s_b^2 = s_\varepsilon^2 \cdot \frac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{5.35.}$$

În etapa testării semnificației estimatorilor se are în vedere probabilitatea ca evoluția lor să se încadreze într-un interval în care frecvențele sunt simetrice în jurul mediei, mai mari în partea centrală și descrescătoare spre capetele intervalului, asemănătoare „clopotului lui Gauss". Această probabilitate este dată de funcția de repartiție Fx. Densitatea de repartiție pentru funcția normală redusă (cu media 0 și dispersia 1) este dată de relația:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \tag{5.36.}$$

Scanned with OKEN Scanner

unde:

$$t = \frac{x_i - \bar{x}}{\sigma_x} \qquad (5.37.)$$

Am folosit variabila t, corespunzătoare repartiției Student, deoarece, în general, eşantioanele sunt de volum redus ($n < 30$). În cazul eşantioanelor de volum mare ($n > 30$) se utilizează variabila z, corespunzătoare repartiției Gauss-Laplace. De regulă, dispersia (σ_x^2), respectiv abaterea medie pătratică (σ_x) a variabilei X nu este cunoscută la nivelul colectivității totale, motiv pentru care ea se aproximează prin dispersia (s_x^2), respectiv abaterea medie pătratică (s_x), la nivelul eşantionului.

În aplicarea procesului de testare a semnificației statistice a parametrilor modelului, se parcurg următoarele etape: se stabileşte ipoteza nulă şi ipoteza alternativă; se stabileşte repartiția în baza căreia se efectuează testarea; se calculează valorea testului statistic (a statisticii) al variabilei standardizate în raport cu mărimea căreia se apreciază calitatea estimării; se fixează valoarea tabelată a variabilei standardizate, corespunzătoare repartiției stabilite; se compară valoarea calculată a statisticii cu cea tabelată. În funcție de raportul (relația de ordine) dintre aceste două mărimi, se acceptă ipoteza nulă (H_0), dacă nivelul calculat al statisticii este mai mic sau egal cu cel tabelat şi respectiv se respinge ipoteza nulă (acceptând-o implicit ca adevărată pe cea alternativă) dacă nivelul calculat al statisticii este mai mare decât cel tabelat.

Dintre coeficienții modelului de regresie, b_0 este panta dreptei şi prezintă o mai mare importanță, deoarece atât timp cât $b_0 \neq 0$ înseamnă că există o legătură liniară directă sau inversă, iar dacă $b_0 = 0$ atunci nu există legătură liniară între variabilele analizate. Aşadar, vom testa dacă b_0 este semnificativ diferit de zero sau nu. Cele două ipoteze vor fi:

H_0: b_0 nu diferă semnificativ de 0 ($b_0 = 0$);

H_1: b_0 diferă semnificativ de 0 ($b_0 \neq 0$).

Dacă vom avea un eşantion de volum mare ($n > 30$), se utilizează testul Z:

$$z_{calc} = \frac{b - \mu(b)}{s_b} = \frac{b - 0}{s_b} \qquad (5.38.)$$

($\mu(b) = b_0$, dar conform ipotezei nule, $b_0 = 0$).

Pentru un prag de semnificație α, se găseşte valoarea tabelară a statisticii din tabelul repartiției Gauss-Laplace şi se compară cu valoarea calculată a testului. Dacă $|z_{calc}| \leq z_{\alpha/2}$, atunci se acceptă ipoteza nulă, conform căreia estimatorul b al

Scanned with OKEN Scanner

lui b_0 provine dintr-o populație cu $b_0 = 0$. Dacă $z_{calc} \leq -z_{\alpha/2}$ sau $z_{calc} \geq z_{\alpha/2}$, se respinge ipoteza nulă, se acceptă ipoteza alternativă, care susține că estimatorul b provine dintr-o populație cu $b_0 \neq 0$.

Dacă volumul eșantionului este mic ($n > 30$), atunci se utilizează testul t, (Student), de forma:

$$t_{calc} = \frac{b - \mu(b)}{s_b} = \frac{b - 0}{s_b} = \frac{b}{s_b} \qquad (5.39.)$$

care urmează o repartiție Student cu $(n - 2)$ grade de libertate.

Se compară nivelul calculat al testului cu nivelul tabelat $t_{\alpha/2,n-2}$, preluat din tabelul repartiției Student pentru un prag de semnificație α și $(n - 2)$ grade de libertate. Regula de decizie este următoarea:

- dacă $t_{calc} \leq t_{\alpha/2,n-2}$ se acceptă ipoteza nulă;
- dacă $t_{calc} \leq -t_{\alpha/2,n-2}$ sau $t_{calc} \geq t_{\alpha/2,n-2}$, atunci nu există suficiente motive pentru a accepta ipozea nulă, deci ea este respinsă, cu riscul α de a efectua o eroare și se acceptă ca adevărată ipoteza alternativă, conform căreia estimatorul b diferă semnificativ de zero (și $b_0 \neq 0$).

În final, după testarea semnificației statistice a estimatorilor parametrilor de regresie trebuie estimate valorile parametrilor de regresie la nivelul colectivității generale (b_0) sub forma intervalului de încredere dat de relația:

$$b - t_{\alpha/2,n-2} \cdot s_b \leq b_0 \leq b + t_{\alpha/2,n-2} \cdot s_b \qquad (5.40.)$$

În loc de a testa, pe rând, semnificația parametrilor modelului de regresie, este posibil să testăm în bloc validitatea modelului de regresie ales, altfel spus măsura în care el permite să se reconstituie valorile reale ale variabilei endogene (y_i), prin intermediul valorilor ajustate $\hat{y}_i$. Acest lucru se poate realiza cu ajutorul analizei dispersionale (ANOVA) și a testului Fisher (F), prin care se verifică, totodată, și ipoteza homoscedasticității din cadrul metodei celor mai mici pătrate.

Pentru aplicarea metodei regresiei, se determină două tipuri de medii: media totală ($\bar{y}$) și media condiționată a lui Y de către X ($\hat{y}_i = a + b \cdot x_i$). Ca atare, plecând de la aceste două tipuri de medii, se poate împărți variația (abaterea) totală a variabilei dependente ($y_i - \bar{y}$) într-o componentă (abatere) care nu este explicată de model ($y_i - \hat{y}_i$), și o componentă (abatere) explicată de influența variabilei factoriale X asupra celei rezultative, adică de model ($\hat{y}_i - \bar{y}$):

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \qquad (5.41.)$$

Scanned with OKEN Scanner

- abaterea neexplicată de model, calculată ca abatere a valorilor reale, empirice ale variabilei rezultative față de cele teoretice, ajustate, obținute pe baza modelului este pusă pe seama factorilor aleatori, reziduali, neincluși în model;
- abaterea explicată de model, calculată ca abatere a valorilor ajustate față de medie este datorată influenței factorilor incluși, studiați în model (variabilei factoriale X).

Prin ridicarea la pătrat a fiecărei abateri și însumarea pentru toate unitățile observate, se obține:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \qquad (5.42.)$$

Vom face notațiile:

$S_y=\sum_{i=1}^{n}(y_i-\bar{y})^2$ = *varianța totală*, determinată ca sumă a pătratelor abaterilor totale;

$S_{y/x}=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ = *varianța sistematică* (*factorială*) sau *explicată* (de model), ca sumă a pătratelor abaterilor datorate regresiei;

$S_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ = *varianța reziduală* (*neexplicată*), ca sumă a pătratelor erorilor (reziduurilor).

Rezultă atunci că:

$$S_y=S_{y/x}+S_e \qquad (5.43.)$$

Dacă împărțim variațele la numărul gradelor de libertate, obținem *dispersii corectate*:

$s_y^2=\frac{S_y}{n-1}$ (dispersia corectată totală);

$s_{y/x}^2=\frac{S_{y/x}}{k}$ (dispersia corectată sistematică, unde k este numărul variabilelor independente);

$s_e^2=\frac{S_e}{n-k-1}$ (dispersia corectată reziduală);

Prin raportarea dispersiei corectate sistematice la cea aleatorie, se obține statistica F, bazată pe repartiția Fisher-Snedecor, conform relației:

Scanned with OKEN Scanner

$$F_{calc} = \frac{s^2_{y/x}}{s^2_e} \qquad (5.44.)$$

Se compară valoarea astfel calculată a testului cu valoarea sa critică, tabelară (din tabelul repartiției Fisher-Snedecor) F_{tab}, corespunzătoare nivelului de semnificație α și gradelor de libertate (k) și $(n - k - 1)$ $F_{\alpha;k;n-k-1}$. Dacă $F_{calc} > F_{\alpha;k;n-k-1}$ înseamnă că modelul de regresie este semnificativ, bine ales și că el explică într-o măsură satisfăcătoare variația variabilei dependente. În această ipoteză, modelul este util prognozei. Dacă însă $F_{calc} \leq F_{\alpha;k;n-k-1}$, acceptăm ipoteza că modelul este neconcludent, incluzând parametrii nesemnificativi sau o funcție greșit aleasă.

Dacă modelul s-a dovedit valid, se poate determina în continuare intervalul de încredere pentru valorile ajustate pe baza modelului:

$$\hat{y}_i - t_{\alpha/2;n-k-1} \cdot s_{\hat{y}} \leq y_i \leq \hat{y}_i + t_{\alpha/2;n-k-1} \cdot s_{\hat{y}} \qquad (5.45.)$$

Dacă luăm în considerare perioada t, atunci relația devine:

$$\hat{y}_t - t_{\alpha/2;n-k-1} \cdot s_{\hat{y}} \leq y_t \leq \hat{y}_t + t_{\alpha/2;n-k-1} \cdot s_{\hat{y}} \qquad (5.46.)$$

În prealabil trebuie determinată dispersia valorilor ajustate.

Previziunile (y') efectuate pe baza modelului se vor situa în intervalul:

$$y'_{n+p} - t_{\alpha/2;n-k-1} \cdot s_{y'} \leq y_t \leq y'_{n+p} + t_{\alpha/2;n-k-1} \cdot s_{y'} \qquad (5.47.)$$

unde $s_{y'}$ este abaterea medie pătratică a previziunilor.

Exemplul 5.2. Pentru testarea validității modelului de regresie găsit la exemplul 5.1., utilizăm testul F:

$$F_{calc} = \frac{\sum(\hat{y}_i - \bar{y})^2}{k} : \frac{\sum(y_i - \hat{y}_i)^2}{n-k-1} = \frac{109677}{1} : \frac{16474}{8} = 53{,}26$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2969}{10} = 296{,}9 \approx 297$$

Valoarea tabelară a testului F din tabelul repartiție Fisher-Snedecor, pentru un prag de semnificație de 5% și gradele de libertate (1,8) este $F_{0,05;1;8} = 5{,}32$. Cum $F_{calc} > F_{tab}$, înseamnă că modelul de regresie ales este semnificativ statistic și explică variația profitului în funcție de numărul de comenzi.

Scanned with OKEN Scanner

5.3.1.3. Previzionarea evoluției variabilei rezultative pe baza modelului de regresie liniară simplă

Previzionarea evoluției viitoare a fenomenelor este scopul final al procesului de modelare. Ea reprezintă operația de determinare a valorii posibile pe care o poate lua fenomenul studiat într-o perioadă următoare, în funcție de anumite valori, strict determinate, ale variabilei sau variabilelor factoriale, înlocuite în modelul econometric (previziune condiționată). De aceea putem spune că previzionarea fenomenelor, în cazul utilizării unui model de regresie liniar multifactorial, se bazează pe extrapolarea funcției de regresie, în ipoteza că factorii de influență își vor exercita în același mod ca și până acum acțiunea asupra fenomenului rezultativ previzionat.

Pentru ca previzionarea să se poată efectua în bune condiții și să fie de o bună calitate, este necesar ca funcția să reflecte o relație esențială, iar variabilele factoriale să aibă o influență într-adevăr semnificativă asupra fenomenului complex studiat, factorii principali să fie corect identificați și acțiunea lor asupra variabilei efect să fie comensurată corect. Totuși, întrucât numărul factorilor de influență al unui fenomen analizat este mare, iar predicția se bazează pe un număr mare de elemente (variabile exogene, influențe întârziate, influențe directe sau indirecte etc.) combinate în cadrul modelului, a căror influență asupra variabilei efect nu poate fi întotdeauna anticipată, putem spune că previziunea nu se poate efectua niciodată cu certitudine, ci este însoțită de un anumit grad de risc, de posibilitatea comiterii unor erori. Totuși, dacă se identifică corect factorii cu acțiune semnificativă asupra fenomenului analizat și dacă se exprimă corect cantitativ această acțiune, modelele elaborate pot permite obținerea unor valori teoretice previzionate destul de apropiate de valorile reale, facilitând cunoașterea mai detaliată a fenomenului complex studiat. Acest lucru este valabil dacă fenomenele se vor desfășura în condiții normale, asigurând reducerea erorilor de previziune și dacă factorii, condițiile care au acționat în orizontul analizat se mențin și în orizontul de prognoză. Totodată, pentru sesizarea corectă a tuturor factorilor de influență care acționează asupra fenomenului complex studiat, este necesar ca perioada pe care se face analiza să fie mai mare, seriile cronologice să aibă un număr suficient de mare de termeni (minim 15 termeni). Orizontul de prognoză, în schimb, trebuie să nu aibă o lungime prea mare, recomandându-se, în practică, ca el să aibă o lungime maximă egală cu o treime din lungimea intervalului pentru care s-a determinat modelul de regresie.

Presupunând că variabila independentă ia valoarea specificată x_p și că legătura liniară simplă descrisă de model se menține, atunci valoarea corespunzătoare a variabilei dependente va fi:

$$Y_p = a_0 + b_0 \cdot x_p + \varepsilon_p \qquad (5.48.)$$

Scanned with OKEN Scanner

Valoarea astfel determinată va diferi de valoarea reală, empirică, datorită acțiunii factorilor perturbatori, reziduali și erorilor de estimare specifice eșantionării.

Diferența dintre valoarea previzionată a variabilei dependente și la momentul/perioada de timp t (Y_t^p) și valoarea reală a fenomenului pentru acea perioadă/moment de timp poartă numele de *eroare de previziune*:

$$\varepsilon_t^p = Y_t^p - y_t \qquad (5.49.)$$

Eroarea de previziune se poate datora unor cauze multiple, precum: acțiunea factorilor aleatorii, reziduali, care nu poate fi controlată și care determină abateri aleatorii ce urmează, de regulă, o repartiție normală; apariția erorilor de estimație a parametrilor, în urma extinderii rezultatelor de la nivelul unui eșantion de volum redus, ce nu provine dintr-o extragere aleatorie (în cazul seriilor cronologice), la nivelul colectivității totale; nu au fost corect identificați toți factorii cu acțiune semnificativă sau acțiunea lor asupra variabilei complexe este descrisă printr-un model neadecvat ales; fixarea unor valori neconforme cu realitatea ale variabilelor exogene pentru perioada pentru care se efectuează prognoza.

Valoarea reală a variabilei previzionate va fi:

$$a_0 + b_0 \cdot x_p - t_{\alpha/2;n-2} \cdot \sigma_p < Y_p < a_0 + b_0 \cdot x_p + t_{\alpha/2;n-2} \cdot \sigma_p \qquad (5.50.)$$

unde σ_p este aproximată prin:

$$s_p = s_\varepsilon \sqrt{\frac{1}{n} + \sum_{i=1}^{n} \frac{\left(x_{ip} - \bar{x}\right)^2}{\left(x_i - \bar{x}\right)^2}} \qquad (5.51.)$$

unde s_ε este abaterea medie pătratică a caracteristicii Y față de Y_p.

Din relația de calcul (5.51.) se observă că eroarea de previziune este invers proporțională cu volumul eșantionului și cu suma pătratelor abaterilor valorilor variabilei factoriale de la media sa. De asemenea, eroarea de previziune crește cu cât abaterea valorii stabilite pentru perioada viitoare (x_{ip}) de la media variabilei factoriale, în perioada în care s-a făcut estimarea este mai mare.

Pentru a testa semnificația previziunii efectuate, se calculează statistica:

$$t_{calc} = \frac{y_p - Y_p}{s_p} \sqrt{n-k} \qquad (5.52.)$$

unde:

y_p este valoarea reală așteptată;
Y_p este valoarea previzionată rezultată din estimare pe baza ecuației de regresie;
k este numărul variabilelor independente incluse în model ($k = 1$).

Scanned with OKEN Scanner

Dacă $t_{calc} \leq t_{\alpha/2,n-2}$, atunci eroarea de prognoză (adică diferența dintre valoarea estimată pe baza modelului și valoarea reală a lui Y) este cu totul întâmplătoare. Dacă, dimpotrivă, $t_{calc} > t_{\alpha/2,n-2}$, erorile de previziune sunt sistematice și deci modelul estimat pentru o perioadă viitoare nu se potrivește cu evoluția reală a fenomenului.

Pentru a obține previziuni mai precise, se poate testa calitatea prognozelor determinate pe baza modelului elaborat, comparându-le cu așa-numitele *prognoze ex-post*[2]. Acestea din urmă sunt valori previzionate, calculate pentru perioade la care se cunosc date reale, putându-se stabili, prin compararea primelor cu cele din urmă, calitatea și precizia prognozei. Pentru elaborarea prognozelor „ex-post" se alege un segment „martor" de timp, care reprezintă, de fapt, un subsegment din orizontul de timp analizat (de regulă ultimele m perioade din intervalul studiat). Pe baza modelului econometric ales se determină valorile estimate ale variabilei endogene dependente, pentru primele $n - m$ perioade/momente de timp (unde n reprezintă numărul total de perioade/momente de timp din intervalul analizat); se fixează, apoi, nivelurile variabilelor exogene independente pentru următoarele m perioade/momente temporale (care sunt și ultimele din orizontul analizat) și se înlocuiesc în model; se determină, astfel, nivelurile previzionate ale variabilei dependente pentru ultimele m perioade sau momente de timp și se compară cu valorile reale, existente. Chiar dacă modelul este bine specificat, neafectat de perturbații sistematice, vor exista totuși erori de prognoză datorită eșantionării și acțiunii factorilor accidentali, erori care pot avea distribuția normală a unei variabile aleatorii.

Pentru a elabora previziuni de bună calitate cu ajutorul unor modele stabilite anterior, trebuie parcursă următoarea succesiune de pași[2]:

I. Efectuarea unor operații pregătitoare elaborării prognozei, de verificare și prelucrare, care să vizeze: eliminarea erorilor sistematice de culgere și înregistrare a datelor; asigurarea omogenității datelor privind unitatea de măsură, forma de exprimare, conținutul, sfera de cuprindere a indicatorilor etc.; utilizarea testelor statistice pentru verificarea validității, a semnificației statistice a modelului elaborat; fixarea nivelului variabilelor exogene pentru orizontul de prognoză etc.

II. Efectuarea unor previziuni „ex-post", pe baza modelului, pentru care cunoaștem valorile reale ale variabilei dependente studiate și stabilirea intervalelor de încredere pentru previziunile obținute; compararea, apoi, a previziunilor „ex-post" cu nivelurile reale ale variabilei dependente studiate, pe intervalul de prognoză și analiza erorilor de prognoză rezultate. Dacă erorile calculate pe segmentul „de probă" („martor") au medie aproximativ nulă și dispersie constantă pe zone ale

---

[2] Pecican E. Șt., *Econometrie*, Editura ALL, București, 1994.

Scanned with OKEN Scanner

intervalului de comparare şi nu sunt autocorelate, atunci se poate considera că modelul este „apt" pentru obţinerea unor valori previzionate apropiate de valorile reale ale fenomenului studiat. Dacă acest lucru nu se întâmplă şi media şi dispersia erorilor de prognoză nu îndeplinesc condiţiile menţionate, se determină motivul, iar dacă nu este găsit, se poate lua în considerare un element de ajustare pentru valorile prognozate, astfel încât evoluţia previzionată să fie atenuată sau amplificată, penturu a se apropia de evoluţia reală a fenomenului analizat.

III. Studiul prognozelor elaborate sub aspect calitativ. În cadrul acestei etape, este necesar să lămurim în ce măsură evoluţia viitoare previzionată este în concordanţă cu teoriile economice referitoare la legităţile care guvernează fenomenele şi procesele economico-sociale. De asemenea, se pot analiza toate elementele ce compun conjunctura economică şi care şi-ar putea pune amprenta asupra evoluţiei viitoare a fenomenului analizat (inclusiv unele aspecte calitative). Pe lângă toate acestea, se pot confrunta şi părerile unor specialişti în viitorologie, privind predicţiile „ex-post" elaborate şi etapa anterioară, precum şi valorile pe care le-ar putea lua variabilele exogene în perioada de prognoză.

IV. Se compară diferitele variante de prognoză elaborate în etapa II, ţinându-se seama şi de aspectele reieşite din analiza calitativă din etapa III, se adoptă o variantă de prognoză care are şansele să corespundă cât mai mult cu realitatea. Pe lângă varianta de bază, se mai pot stabili şi alte variante adiacente, alternative, care se centrează pe scenarii mai optimiste sau mai pesimiste.

V. Actualizarea modelului în funcţie de schimbările ce pot apărea în conjunctura economică, introducerea de noi date sau de noi variabile, dacă este cazul, verificarea validităţii relaţiilor între variabilele economice, descrise prin ecuaţiile modelului etc.

5.3.1.4. Analiza statistică a factorilor aleatorii

Am văzut că fenomenele complexe sunt influenţate de un număr mare de factori de influenţă, ce pot fi împărţiţi în două categorii: factori sistematici (a căror acţiune asupra variabilei complexe este studiată în cadrul modelului) şi factori aleatorii, reziduali (adică toţi ceilalţi factori, care n-au fost incluşi în model, care îşi pun amprenta asupra fenomenului efect şi care sunt exprimaţi printr-o variabilă reziduală). În această secţiune ne concentrăm atenţia asupra factorilor din această ultimă categorie.

Factorii reziduali sunt mai dificil de măsurat, de exprimat numeric, valorile variabilei reziduale rezultând prin compararea valorilor ajustate, teoretice, calculate pe baza modelului elaborat cu valorile reale, empirice ale variabilei dependente. Existenţa factorilor aleatorii, neincluşi în model duce la posibilitatea de apariţie a unor erori, care se referă la: neidentificarea corectă a principalilor factori (sau a unui număr insuficient de factori) ce ar putea influenţa variabila rezultativă

Scanned with OKEN Scanner

studiată; identificarea greşită a formei legăturii dintre variabilele analizate; cuantificarea eronată a acţiunii unor fenomene (în special a fenomenelor calitative, care sunt mai dificil de exprimat numeric); greşeli survenite în formarea eşantionului.

După ce fenomenul analizat a fost descris cu ajutorul modelului, rezultă şi valorile reziduale, indicând în ce măsură modelul (ecuaţia sau ecuaţiile) adoptat descrie desfăşurarea fenomenului în funcţie de factorii pe care i-am considerat importanţi. Ipotezele referitoare la eventualele perturbaţii pot fi considerate adevărate „etaloane", „praguri" sau „baremuri" în raport cu care se va aprecia calitatea modelului şi calitatea soluţiilor obţinute pe baza acestuia, prin metoda de estimare aleasă. Pentru aprecierea valorilor variabilei rezultative se iau în considerare ipotezele Metodei Celor Mai Mici Pătrate (MCMMP), verificându-se dacă valorile variabilei perturbatoare (aleatorii) (ε_{i0} – la nivelul colectivităţii totale şi, respectiv, ε_i la nivelul eşantionului) îndeplinesc aceste ipoteze. În cazul nerespectării, este bine să se identifice ce consecinţe ar putea avea acest fapt asupra calităţii modelului, precum şi modalităţile de eliminare sau de atenuare, pe cât posibil, a acestor consecinţe.

Valorile reziduale sunt definite de relaţiile:

$$\varepsilon_i = y_i - \hat{y}_i \text{ (la nivelul eşantionului)} \qquad (5.53.)$$

ele fiind o măsură a variaţiei valorilor caracteristicii Y în jurul dreptei de regresie (abaterile valorilor ajustate faţă de cele reale ale variabilei Y).

$$\varepsilon_{i0} = y_i - a_0 - b_0 \cdot x_i \text{ (la nivelul colectivităţii totale)} \qquad (5.54.)$$

Pentru aplicarea metodei regresiei, am văzut că se fac următoarele presupuneri asupra erorilor reziduale ε_i (respectiv ε_{i0}):

- erorile reziduale sunt variabile aleatorii independente;
- erorile reziduale provin din populaţii normal distribuite;
- media erorilor reziduale este nulă;
- dispersia erorilor reziduale este constantă, oricare ar fi valorile lui X.

Dacă nu sunt respectate aceste ipoteze, nu se poate aplica metoda regresiei liniare astfel încât să obţinem rezultate de calitate.

De exemplu, am văzut că una din problemele ce pot apărea atunci când construim un model de regresie este autocorelaţia. Într-adevăr, se presupune iniţial că valorile variabilei reziduale sunt independente, nu sunt autocorelate. Totuşi, deseori, această ipoteză este infirmată, din motive precum: analizăm o serie cronologică ce se referă la o variabilă endogenă, dependentă, a cărei evoluţie la un moment dat este influenţată de nivelul şi evoluţia dintr-o perioadă trecută, lucru ce poate genera în timp autocorelarea reziduurilor; s-a omis introducerea unei variabile independente, exogene, cu influenţă semnificativă asupra variabilei rezultative; s-au folosit date care se referă la perioade mai mici de un an (trimestriale sau lunare), care sunt afectate de sezonalitate, sezonalitate care se va

Scanned with OKEN Scanner

manifesta similar de-a lungul unor ani succesivi, ceea ce înseamnă că va exista o anumită corelație între termenii reziduali.

Testul cel mai important și mai des folosit pentru verificarea autocorelării termenilor reziduali este testul Durbin-Watson, care constă în calcularea statisticii d_{calc} cu ajutorul relației:

$$d_{calc} = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=2}^{n}\varepsilon_t^2} = 2\left(1 - r_{\varepsilon_t,\varepsilon_{t-1}}\right) \qquad (5.55.)$$

În cazul acestui test, ipotezele nulă și alternativă sunt:

H_0: valorile variabilei reziduale nu sunt autocorelate ($\rho = 0$);

H_1: valorile variabilei reziduale sunt pozitiv autocorelate ($\rho > 0$);

sau

H_1: valorile variabilei reziduale sunt negativ autocorelate ($\rho < 0$);

(unde ρ = valoarea coeficientului sau raportului de corelație la nivelul întregii colectivități).

Se observă că testul este întotdeauna unilateral, motiv pentru care nu putem testa decât fie ipoteza autocorelației pozitive, fie ipoteza autocorelației negative, dar niciodată ambele simultan.

Din relația (5.55.) se observă că definirea testului Durbin-Watson cuprinde în expresia sa coeficientul de autocorelație la nivelul eșantionului r, motiv pentru care domeniul de valori al testului depinde de intervalul în care ia valori coeficientul de autocorelație ([–1, 1]). Astfel, în cazul autocorelației perfecte pozitive, $r_{\varepsilon_t,\varepsilon_{t-1}} = +1$ și deci $d_{calc}^{\min} = 0$; în cazul existenței unei autocorelații perfect negative între valorile variabilei reziduale, $r_{\varepsilon_t,\varepsilon_{t-1}} = -1$ și deci $d_{cal}^{\max} = 4$. Rezultă, așadar, că testul Durbin-Watson poate lua valori între 0 și 4, cu o valoare „ideală" (2), indicând absența autocorelației (caz în care $r_{e_t,e_{t-1}} = 0$).

Valoarea calculată a testului se compară cu valori critice, tabelate, corespunzătoare unui prag de semnificație α (de regulă se alege $\alpha = 5\%$), numărului de observații n și numărului de variabile independente k ($d_{\alpha,n,k}$). De fapt, sunt preluate din tabel două valori limită (o valoare minimă: d_{inf} și o valoarea maximă: d_{sup}), generate în ipoteza repartiției normale, pentru diverse praguri de semnificație. În diagrama 5.6. și în Tabelul 5.4. se prezintă modul în care se adoptă deciziile (nulă sau alternativă):

Scanned with OKEN Scanner

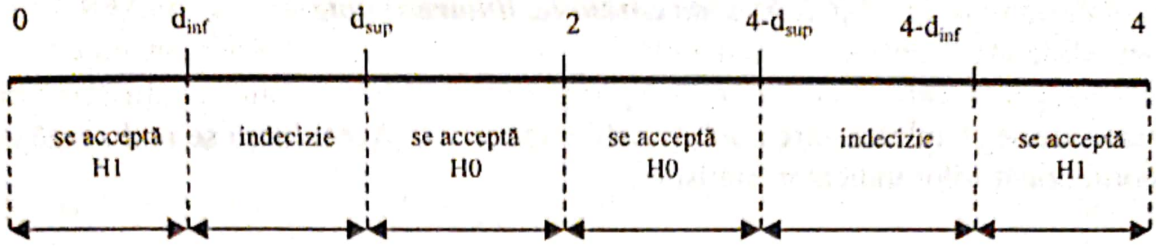


Figura 5.6. Regulile de adoptare a deciziilor în testul Durbin-Watson

Regulile de adoptare a deciziilor în testul Durbin-Watson

Tabelul 5.4.

Intervale	Decizia acceptată	Comentariu
$d_{calc} = 0$	H_1	Autocorelare pozitivă perfectă
$d_{calc} < d_{inf}$	H_1	Autocorelare pozitivă
$d_{inf} < d_{calc} < d_{sup}$	Nici o decizie	Rezultat neconcludent
$d_{sup} < d_{calc} < 4 - d_{sup}$	H_0	Nu există autocorelație
$4 - d_{sup} < d_{calc} < 4 - d_{inf}$	Nici o decizie	Rezultat neconcludent
$d_{calc} > 4 - d_{inf}$	H_1	Autocorelare negativă
$d_{calc} = 4$	H_1	Autocorelare negativă perfectă

Se observă că, dacă valoarea calculată a testului este mai mică decât valoarea critică inferioară ($d_{calc} < d_{inf}$), atunci respingem ipoteza nulă şi o considerăm adevărată pe cea alternativă, concluzionând că există o autocorelare pozitivă. Dacă valoarea calculată a testului este mai mare decât valoarea critică superioară ($d_{calc} > d_{sup}$) atunci se adoptă ipoteza nulă, conform căreia valorile variabilei reziduale nu sunt autocorelate. Dacă valoarea calculată a testului este situată între cele două limite critice ($d_{inf} < d_{calc} < d_{sup}$), spunem că rezultatul testului este incert şi pe baza lui nu se poate adopta nici o decizie. Asemănător se procedează pentru testarea autocorelării negative a valorilor variabilei reziduale, pentru identificarea celor două valori critice limită utilizându-se transformarea: $d = 4 - d$.

Scanned with OKEN Scanner

5.3.2. Metoda corelației liniare simple

Metoda corelației liniare simple permite măsurarea intensității legăturii liniare care se stabilește între două variabile numerice. Acest lucru se realizează cu ajutorul următorilor indicatori statistici:

a) covarianța

Să presupunem că, pentru cele n unități statistice ale unei populații, s-au înregistrat variabilele X și Y, cu valorile (x_i, y_i) pentru unitatea statistică i, $i = \overline{1,n}$, pentru care s-au calculat mediile $\bar{x}$ și $\bar{y}$. Dacă împărțim planul delimitat de cele două axe Ox și Oy din graficul de corelație în cadrane, în funcție de cele două valori medii, atunci acesta va arăta ca în Figura 5.7.

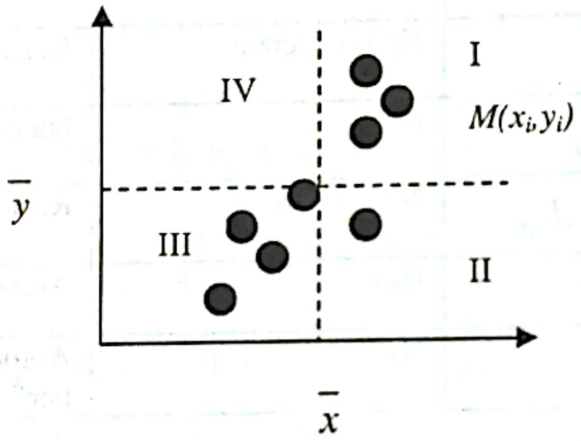


Figura 5.7. Corelograma

Pentru abaterile valorilor individuale ale fiecărei caracteristici față de media lor, de forma: $d_i^x = x_i - \bar{x}$ și $d_i^y = y_i - \bar{y}$ se studiază semnul acestora, precum și semnul produselor abaterilor, în fiecare din cele patru cadrane definite. Rezultatele sunt prezentate în Tabelul 5.5.

Tabelul 5.5.

Cadranul	Semnul pentru:		
	d_i^x	d_i^y	$d_i^x \cdot d_i^y = (x_i - \bar{x})(y_i - \bar{y})$
I	+	+	+
II	+	–	–
III	–	–	+
IV	–	+	–

Scanned with OKEN Scanner

Rezultă că, dacă majoritatea punctelor din corelogramă, de coordonate (x_i, y_i) sunt amplasate în cadranele I şi III, atunci legătura dintre variabile este directă, iar dacă cele mai multe dintre punctele corelogramei se regăsesc în cadranele II şi IV, atunci legătura va fi inversă. Însumând produsele abaterilor d_i^x şi d_i^y şi împărţind suma lor la n (adică calculând media aritmetică a produselor abaterilor valorilor individuale faţă de media lor), se obţine un indicator numit ***covarianţă***, definit de relaţia:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{n} \qquad (5.56.)$$

Prelucrând relaţia (5.56.), obţinem o formă echivalentă a covarianţei:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}\left(x_i y_i-x_i\cdot\bar{y}-\bar{x}\cdot y_i+\bar{x}\cdot\bar{y}\right)}{n}=\frac{\sum_{i=1}^{n}x_i y_i}{n}-\bar{y}\cdot\frac{\sum_{i=1}^{n}x_i}{n}-\bar{x}\cdot\frac{\sum_{i=1}^{n}y_i}{n}+\frac{n\cdot\bar{x}\cdot\bar{y}}{n}=$$

$$=\frac{\sum_{i=1}^{n}x_i y_i}{n}-2\cdot\bar{x}\cdot\bar{y}+\bar{x}\cdot\bar{y}=\overline{xy}-\bar{x}\cdot\bar{y} \qquad (5.57.)$$

Covarianţa se mai poate defini, deci, conform relaţiei (5.57.), ca diferenţa dintre media produselor celor două variabile şi produsul mediilor lor. Ea are o unitate de măsură concretă, dependentă de unităţile de măsură ale celor două variabile şi nu are un interval fix de variaţie, putând lua valori oricât de mari sau oricât de mici (în funcţie de valorile caracteristicilor şi de diferenţele între ele). De aceea covarianţa nu poate fi utilizată la măsurarea intensităţii legăturii, dar semnul ei poate indica sensul legăturii:

- dacă $\operatorname{cov}(x,y)>0$, atunci legătura este directă;
- dacă $\operatorname{cov}(x,y)<0$, atunci legătura este inversă;
- dacă $\operatorname{cov}(x,y)=0$, atunci nu există legătură între variabile (variabile independente).

Din relaţia (5.22.) ştim că $b=\frac{\operatorname{cov}(x,y)}{s_x^2}$, de unde rezultă că

$$\operatorname{cov}(x,y)=b\cdot s_x^2 \qquad (5.58.)$$

Se observă că semnul covarianeţi coincide întotdeauna cu semnul lui b.

Scanned with OKEN Scanner

b) Coeficientul de corelație liniară simplă al lui Pearson

Stabilește limita până la care punctele de coordonate (x_i, y_i) sunt amplasate de-a lungul unei drepte liniare și e dat de raportul între covarianță și produsul abaterilor medii pătratice ale celor două variabile:

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot s_x \cdot s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \qquad (5.59.)$$

Relația (5.59.) se poate rescrie într-o formă echivalentă:

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n \cdot \bar{y}^2\right)}} \qquad (5.60.)$$

sau

$$r_{xy} = \frac{\Delta b}{\sqrt{\Delta\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]}} \qquad (5.61.)$$

O altă relație de calcul a coeficientului de corelație liniară simplă provine din scrierea covarianței în funcție de coeficientul de regresie b, ca în relația (5.58.).

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{s_x \cdot s_y} = b \cdot \frac{s_x^2}{s_x \cdot s_y} = b \cdot \frac{s_x}{s_y} \qquad (5.62.)$$

Deoarece abaterile medii pătratice ale variabilelor (s_x și s_y) sunt pozitive, rezultă de aici că întotdeauna semnul coeficientului de corelație coincide cu semnului lui b. Coeficientul de corelație liniară simplă ia întotdeauna valori în intervalul [–1, +1], el ne indică sensul legăturii, dar măsoară și intensitatea ei.

- dacă $r_{xy} > 0$, există o legătură directă între variabile;
- dacă $r_{xy} < 0$, există o legătură inversă între variabile;
- dacă $r_{xy} = 0$, nu există legătură liniară între variabile;
- dacă $r_{xy} \to \pm 1$, există o legătură puternică între variabile (directă sau inversă);
- dacă $r_{xy} \to 0$, există o legătură slabă între variabile (directă sau inversă).

Scanned with OKEN Scanner

Altfel spus, putem întâlni următoarele situații:

- $r_{xy} = 1$ — Legătură directă de tip determinist; punctele din graficul de corelație se înșiră exact pe o dreaptă cu pantă pozitivă [vezi Figura 5.1. f)];
- $0{,}95 \le r_{xy} < 1$ — Legătură directă foarte puternică; punctele din corelogramă sunt foarte strâns adunate în jurul dreptei de regresie cu pantă pozitivă;
- $0{,}75 \le r_{xy} < 0{,}95$ — Legătură directă puternică [Figura 5.1. a)];
- $0{,}5 \le r_{xy} < 0{,}75$ — Legătură directă de intensitate medie; punctele sunt ceva mai răsfirate în jurul dreptei, dar se remarcă o oarecare concentrare de o parte sau de alta a ei;
- $0{,}2 \le r_{xy} < 0{,}5$ — Legătură directă de intensitate slabă; punctele sunt mai răspândite în planul corelogramei;
- $0 < r_{xy} < 0{,}2$ — Nu există legătură semnificativă între variabile, punctele sunt dispersate în tot planul corelogramei;
- $-0{,}2 < r_{xy} < 0$
- $r_{xy} = 0$ — Absența legăturii; puncte răspândite uniform în plan [Figura 5.1. c)];
- $-0{,}5 < r_{xy} \le -0{,}2$ — Legătură inversă de intensitate slabă;
- $-0{,}75 < r_{xy} \le -0{,}5$ — Legătură inversă de intensitate medie; se remarcă o oarecare concentrare a punctelor din corelogramă într-o zonă orientată pe direcția unei drepte cu pantă negativă;
- $-0{,}95 < r_{xy} \le -0{,}75$ — Legătură inversă puternică; punctele sunt strâns adunate în jurul unei drepte cu pantă negativă [Figura 5.1. b)];
- $-1 < r_{xy} \le -0{,}95$ — Legătură inversă foarte puternică; punctele sunt foarte strâns concentrate în jurul dreptei de regresie cu pantă negativă;
- $r_{xy} = -1$ — Legătură inversă de tip determinist; punctele se înșiră exact pe o dreaptă cu pantă negativă [Figura 5.1. g)].

Dacă fiecare pereche de valori (x_i, y_i) apare de n_i ori, atunci coeficientul de corelație se scrie:

$$r_{xy} = \frac{\sum_{i=1}^{r} n_i \sum_{i=1}^{r} x_i y_i n_i - \sum_{i=1}^{r} x_i n_i \sum_{i=1}^{r} y_i n_i}{\sqrt{\left[\sum_{i=1}^{r} n_i \sum_{i=1}^{r} x_i^2 n_i - \left(\sum_{i=1}^{r} x_i n_i\right)^2\right]\left[\sum_{i=1}^{r} n_i \sum_{i=1}^{r} y_i^2 n_i - \left(\sum_{i=1}^{r} y_i n_i\right)^2\right]}} \qquad (5.63.)$$

Scanned with OKEN Scanner

Dacă datele au fost sistematizate în grupe după variația ambelor caracteristici, într-un tabel cu dublă intrare și fiecare pereche de valori (x_i, y_j) apare de n_{ij} ori ($i = \overline{1,r}$, $j = \overline{1,p}$), atunci coeficientul de corelație va fi:

$$r_{xy} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{p} n_{ij} \sum_{i=1}^{r}\sum_{j=1}^{p} x_i y_i n_{ij} - \sum_{i=1}^{r} x_i n_{i.} \sum_{i=1}^{r}\sum_{j=1}^{p} y_j n_{.j}}{\sqrt{\left[\sum_{i=1}^{r} n_{i.} \sum_{i=1}^{r} x_i^2 n_{i.} - \left(\sum_{i=1}^{r} x_i n_{i.}\right)^2\right]\left[\sum_{j=1}^{p} n_{.j} \sum_{j=1}^{p} y_j^2 n_{.j} - \left(\sum_{j=1}^{p} y_j n_{.j}\right)^2\right]}} \quad (5.64.)$$

Coeficientul de corelație corespunzător populației totale va fi:

$$\rho_{xy} = \frac{COV(x,y)}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} = \frac{COV(x,y)}{\sigma_x \cdot \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad (5.65.)$$

unde σ_x, σ_y sunt abaterile medii pătratice ale celor două variabile la nivelul colectivității totale.

Exemplul 5.3. Pentru datele din Tabelul 5.2., vom determina coeficientul de corelație liniară simplă:

$$r_{xy} = \frac{109833}{\sqrt{11001(10.077.530 - 8.814.961)}} = 0{,}93$$

Între numărul de comenzi și profitul obținut există o legătură directă foarte puternică.

Testarea semnificației coeficientului de corelație

De regulă, datele utilizate în calcularea coeficientului de corelație (r_{xy}) provin dintr-un eșantion al unei populații generale a tuturor observațiilor posibile, motiv pentru care coeficientul de corelație (r_{xy}) este influențat de fluctuațiile de selecție. În cele mai multe cazuri coeficientul de corelație al populației totale (ρ_{xy}) este necunoscut, astfel că el va fi estimat prin intermediul coeficientului de corelație la nivelul eșantionului (r_{xy}). Pentru testarea semnificației indicatorului cu ajutorul ***testului t*** (Student) este necesar să fie îndeplinită condiția asupra normalității distribuțiilor ambelor variabile studiate (X, Y).

Ipoteza nulă și ipoteza alternativă a *testului t* se referă la presupunerea că indicatorul „coeficientul de corelație liniară simplă" la nivelul colectivității generale este semnificativ diferit de zero sau nu:

H_0: $\rho_{xy} = 0$ (5.66.)

H_1: $\rho_{xy} \neq 0$ (5.67.)

Scanned with OKEN Scanner

Se determină statistica t, conform relației:

$$t_{calc} = r_{xy}\sqrt{\frac{n-2}{1-r^2}} \qquad (5.68.)$$

Se compară valoarea astfel calculată a testului cu valoarea tabelară, preluată din tabelul repartiției Student pentru un prag de semnificație α și $n-2$ grade de libertate ($t_{\alpha/2;n-2}$). Regula de decizie este următoarea:

- dacă $t_{calc} \leq t_{\alpha/2;n-2}$, se acceptă ipoteza nulă ca adevărată, deci ρ_{xy} nu diferă semnificativ de zero.
- Dacă $t_{calc} > t_{\alpha/2;n-2}$, se acceptă ipoteza alternativă ca fiind adevărată, deci avem suficiente motive să presupunem că ρ_{xy} diferă semnificativ de zero.

Exemplul 5.4. Vom testa semnificația coeficientului de corelație calculat în exemplul (5.3.).

$$t_{calc} = r_{xy}\sqrt{\frac{n-2}{1-r^2}} = 0{,}93 \cdot \sqrt{\frac{8}{1-0{,}93^2}} = 7{,}156$$

Se compară t_{calc} cu $t_{\alpha/2;n-2} = t_{0{,}025;8} = 2{,}306$

Cum $t_{calc} > t_{\alpha/2;n-2}$, rezultă că indicatorul *coeficientul de corelație* la nivelul colectivității generale este semnificativ statistic (este diferit de zero).

c) Raportul de corelație

Spre deosebire de coeficientul de corelație, care se determină numai în cazul legăturilor liniare între variabile, raportul de corelație se determină pentru măsurarea intensității legăturilor liniare ***și neliniare***.

Raportul de corelație are următoarea relație de calcul:

$$R_{y/x} = \sqrt{1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}} = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} \qquad (5.69.)$$

Întrucât este adevărată egalitatea de descompunere a variației totale a lui Y în funcție de influențele celor două categorii principale de factori (prima categorie este formată din X – factor sistematic și a doua categorie este formată din factorii aleatori, adică ceilalți factori, în afară de X, care ar putea influența variația variabilei rezultative și care nu sunt incluși în studiu), avem relația:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (5.70.)$$

Scanned with OKEN Scanner

Dacă se împarte egalitatea (5.70.) la n, se obține o relație între dispersii:

$$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{n}+\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}. \qquad (5.71.)$$

unde:

$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=s_y^2$ reprezintă *dispersia totală* a variabilei rezultative, dată de influența tuturor factorilor sistematici și aleatorii asupra lui Y.

$\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{n}=s_{y/x}^2$ reprezintă *dispersia sistematică*, ce reflectă variația lui Y sub influența factorului sistematic X (adică a modelului de regresie).

$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}=s_\varepsilon^2$ reprezintă *dispersia reziduurilor*, ce reflectă variația lui Y datorată influenței factorilor aleatorii.

Așadar, raportul de corelație se poate scrie și pe baza dispersiilor:

$$R_{y/x}=\sqrt{1-\frac{s_\varepsilon^2}{s_y^2}}=\sqrt{\frac{s_{y/x}^2}{s_y^2}} \qquad (5.72.)$$

Raportul de corelație ia valori în intervalul [0,1]. Având doar valori pozitive, el nu poate indica direcția (sensul legăturii), ci doar intensitatea ei. Astfel, putem întâlni următoarele situații:

- dacă $R_{y/x}=0$, atunci nu există legătură între variabile;
- dacă $R_{y/x}\to 0$ (în sensul că are valori apropiate de zero), atunci există o legătură slabă între variabile;
- dacă $R_{y/x}\to 1$ (în sensul că are valori apropiate de 1), atunci există o legătură puternică între variabile;
- dacă $R_{y/x}=1$, atunci legătura dintre variabile este foarte puternică, de tip determinist.

În cazul unei legături liniare, este adevărată relația:

$$R_{y/x}\cong|r_{xy}| \qquad (5.73.)$$

Scanned with OKEN Scanner

Egalitatea (5.73.) poate servi şi drept criteriu pentru confirmarea existenţei sau inexistenţei unei legături liniare.

Pătratul raportului de corelaţie $R^2_{y/x}$ se numeşte ***raport de determinaţie***, care, exprimat în procente, arată cât la sută din variaţia totală a lui Y este determinată de modelul de regresie (adică de influenţa lui X).

Exemplul 5.5. Pentru datele din Tabelul 5.2. determinăm raportul de corelaţie:

$$R_{y/x} = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} = \sqrt{\frac{109.677}{126.257}} = \sqrt{0,87} = 0,932 \cong |r_{xy}|$$

Calculele intermediare se găsesc în Tabelul 5.3.

$R^2_{y/x} = 0,87$, deci 87% din variaţia totală a profitului este determinată de modificarea numărului comenzilor. Cum $R_{y/x} \cong |r_{xy}|$, înseamnă că legătura dintre cele două variabile este de tip liniar.

5.3.3. Regresia şi corelaţia neliniară simplă

Deseori, relaţia dintre una sau mai multe variabile cauzale şi o variabilă rezultativă nu se rezumă la o legătură liniară, ci necesită apelarea la o funcţie de regresie neliniară. Alegerea tipului adecvat al funcţiei de regresie se face pe baza graficului de corelaţie. Să presupunem că avem o variabilă factorială şi una rezultativă (regresie simplă). Principalele tipuri de funcţii de regresie neliniară sunt:

a) funcţia parabolică (parabola de gradul doi)

Se alege pentru modelarea tendinţei legăturii dintre două variabile, atunci când această tendinţă are un singur punct de inflexiune (Figura 5.8. a).

Scanned with OKEN Scanner

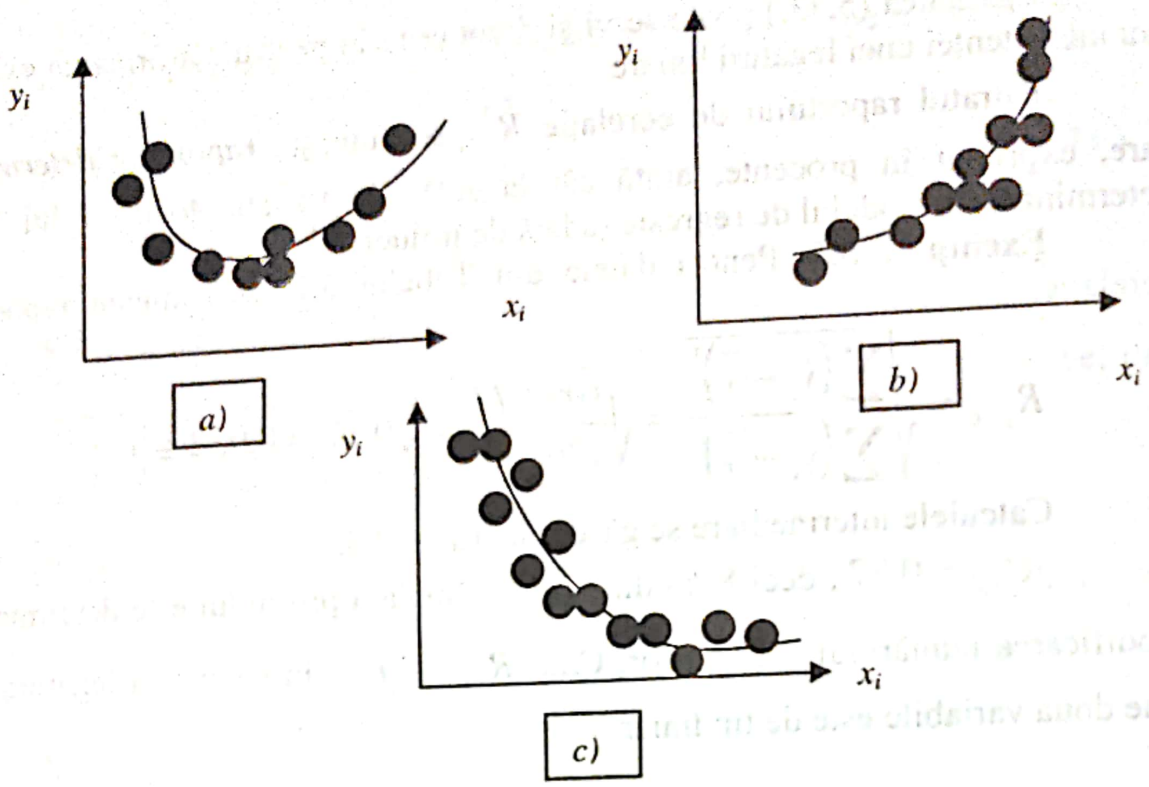


Figura 5.8. Tipuri de legături neliniare

Ecuația de regresie va fi:

$$y_i = a + b \cdot x_i + c \cdot x_i^2 + \varepsilon_i \qquad (5.74.)$$

$$\hat{y}_i = a + b \cdot x_i + c \cdot x_i^2 \qquad (5.75.)$$

Prin aplicarea metodei celor mai mici pătrate se obține un sistem de trei ecuații cu trei necunoscute:

$$\begin{cases} na + b\sum x_i + c\sum x_i^2 = \sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum x_i^2 y_i \end{cases} \qquad (5.76.)$$

Din sistem se determină parametrii a, b, c, şi apoi valorile ajustate ale lui Y.

b) funcția exponențială [Figura 5.8. b)].

Ecuația de regresie este:

$$\hat{y}_i = a \cdot b^{xi} \qquad (5.77.)$$

Prin logaritmare, relația se liniarizează:

$$\lg y_i = \lg a + x_i \lg b \qquad (5.78.)$$

Scanned with OKEN Scanner

Relația (5.78.) exprimă o legătură liniară între logaritmul variabilei Y și variabila X. Prin aplicarea metodei celor mai mici pătrate, se obține un sistem de două ecuații cu necunoscutele $\lg a$ și $\lg b$:

$$\begin{cases} n \cdot \lg a + (\lg b)\sum x_i = \sum \lg y_i \\ (\lg a)\sum x_i + (\lg b)\sum x_i^2 = \sum x_i \lg y_i \end{cases} \qquad (5.79.)$$

Se determină $\lg a$ și $\lg b$ apoi, prin antilogaritmare, parametrii a și b și valorile ajustate ale lui Y în funcție de X. Făcând notațiile $\lg y_i = y_i'$, $\lg a = a'$ și $\lg b = b'$ în ecuația (5.78), se ajunge la modelul liniar prezentat anterior.

c) funcția hiperbolică [Figura 5.8. c)]

Se utilizează atunci când între variabila cauzală și cea rezultativă există o relație de inversă proporționalitate, redată prin modelul:

$$\hat{y}_i = a + b \cdot \frac{1}{x_i} \qquad (5.80.)$$

În urma aplicării metodei celor mai mici pătrate, se obține următorul sistem de două ecuații cu două necunoscute:

$$\begin{cases} na + b\sum \frac{1}{x_i} = \sum y_i \\ a\sum \frac{1}{x_i} + b\sum \frac{1}{x_i^2} = \sum \frac{1}{x_i} y_i \end{cases} \qquad (5.81.)$$

Se determină parametrii a și b, apoi valorile ajustate. Făcând notația $\frac{1}{x_i} = x_i'$ în relația (5.80.), se ajunge la modelul liniar de regresie simplă.

d) funcția polinomială de ordin superior

Dacă se observă că tendința legăturii dintre variabile are mai multe puncte de inflexiune, atunci se poate folosi pentru modelarea acestei tendințe un model polinomial de ordin superior, de forma:

$$\hat{y}_i = a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_k x_i^k \qquad (5.82.)$$

unde k este gradul funcției. Se observă că pentru $k = 2$ se ajunge la modelul parabolic prezentat anterior. În general, regresia polinomială este văzută și studiată ca un caz special de regresie multiplă.

Scanned with OKEN Scanner

Pentru măsurarea intensității legăturii neliniare se poate folosi raportul de corelație, în forma definită prin relația (5.69.) sau (5.72). Pătratul raportului de corelație (raportul de determinație) are aceeași semnificație ca cea dată anterior.

În plus, dacă vrem să determinăm calitatea ajustării, putem utiliza abaterea medie pătratică a reziduurilor, ca și la legătura liniară [vezi relația (5.30.)].

Cel mai potrivit model de regresie utilizat în ajustarea unei variabile rezultative în funcție de valorile fixate ale unei variabile cauzale va fi acela care asigură valoarea minimă a sumei pătratelor abaterilor între valorile reale și cele ajustate ale variabilei dependente Y:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \rightarrow \text{minim} \qquad (5.83.)$$

5.3.4. Regresia și corelația liniară multiplă

Pentru stabilirea formei legăturii dintre variabile, utilizăm una din metodele elementare de analiză a dependențelor între fenomenele economico-sociale: metoda grafică (a corelogramei, a norului de puncte sau a diagramei de împrăștiere), pe baza căreia alegem acea funcție care descrie cel mai bine, cel mai fidel, „linia" prefigurată de norul de puncte trasat pe diagramă. Întrucât vor fi luate în studiu mai multe variabile factoriale care ar putea exercita o influență demnă de luat în seamă asupra variabilei rezultative, înseamnă că ne situăm în sfera regresiei multiple, caz în care norul de puncte va determina un plan, ce poate fi descris printr-o diagramă de împrăștiere tridimensională. Dacă reprezentarea grafică a dependențelor dintre variabilele luate în studiu se orientează în jurul unei drepte din planul descris, atunci folosim un model de regresie liniar multifactorial, la nivelul colectivității generale, dat de ecuația:

$$Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \ldots + \alpha_k X_{ik} + \varepsilon_{0i} \qquad (5.84.)$$

pentru: $i = \overline{1, N}$ reprezentând numărul unității statistice observate din colectivitatea generală;

Y_i = nivelul variabilei complexe, dependente, la unitatea de observare i din colectivitatea generală;

$X_{i1}, X_{i2}, \ldots, X_{ik}$ = valorile celor k variabile factoriale, înregistrate la unitatea statistică i;

$\alpha_0, \alpha_1, \ldots, \alpha_k$ = parametrii modelului la nivelul colectivității generale, ce vor trebui estimați;

ε_{0i} = nivelul variabilei reziduale la unitatea statistică i.

La nivelul eșantionului, se poate scrie:

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_k x_{ik} + \varepsilon_i \qquad i = \overline{1, n} \qquad (5.85.)$$

cu componenta predictibilă:

Scanned with OKEN Scanner

$$\hat{y}_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + ... + a_k x_{ik} \quad i = \overline{1,n} \qquad (5.86.)$$

De menţionat că la ipotezele modelului de regresie liniară simplă, arătate anterior, în cazul modelului de regresie liniară multiplă, se mai adaugă încă una, aceea că oricare din cele k variabile independente, factoriale, nu poate fi scrisă ca o combinaţie liniară perfectă a celorlalte variabile independente (numită *ipoteza necoliniarităţii variabilelor factoriale*). Cu alte cuvinte, nu există un şir de numere $c_0, c_1, ..., c_k$, astfel încât:

$$c_0 + c_1 x_{i1} + c_2 x_{i2} + ... + c_k x_{ik} = 0 \quad i = \overline{1,n} \qquad (5.87.)$$

Parametrii modelului de regresie liniară multiplă de la nivelul colectivităţii generale se determină ca estimatori, pe baza datelor de la nivelul eşantionului. Pentru determinarea estimatorilor parametrilor modelului se utilizează metoda celor mai mici pătrate, care presupune minimizarea sumei pătratelor abaterilor dintre valorile reale, empirice (y_i) şi valorile teoretice, ajustate, rezultate pe baza modelului ($\hat{y}_i$) [relaţia (5.83)].

Estimarea va fi cu atât mai ridicată calitativ, cu cât costul aplicării metodei de estimare este minim şi estimările sunt nedistorsionate (nedeplasate), consistente, eficiente. Dacă componenta predictibilă a modelului de regresie liniară multiplă la nivelul eşantionului este:

$$\hat{y}_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + ... + a_k x_{ik} \quad i = \overline{1,n} \qquad (5.88.)$$

atunci metoda celor mai mici pătrate se scrie

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + ... + a_k x_{ik})]^2 \to \text{minim}, \; i = \overline{1,n} \qquad (5.89.)$$

În urma derivării şi anulării derivatelor parţiale în raport cu a_i ($\frac{\partial S}{\partial a_i} = 0$, $i = \overline{1,k}$) se obţine un sistem de (k + 1) ecuaţii normale cu (k + 1) necunoscute, de forma:

Scanned with OKEN Scanner

$$\begin{cases} na_0 + a_1\sum_i x_{i1} + a_2\sum_i x_{i2} + \ldots + a_k\sum_i x_{ik} = \sum_i y_i \\ a_0\sum_i x_{i1} + a_1\sum_i x_{i1}^2 + a_2\sum_i x_{i1}\cdot x_{i2} + \ldots + a_k\sum_i x_{i1}\cdot x_{ik} = \sum_i x_{i1}\cdot y_i \\ a_0\sum_i x_{i2} + a_1\sum_i x_{i1}\cdot x_{i2} + a_2\sum_i x_{i2}^2 + \ldots + a_k\sum_i x_{i2}\cdot x_{ik} = \sum_i x_{i2}\cdot y_i \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0\sum_i x_{ik} + a_1\sum_i x_{i1}\cdot x_{ik} + a_2\sum_i x_{i2}\cdot x_{ik} + \ldots + a_k\sum_i x_{ik}^2 = \sum_i x_{ik}\cdot y_i \end{cases} \quad (5.90.)$$

Pentru fiecare din cele n unități de observare din eșantion, au loc egalitățile:

$$\begin{aligned} &i=1 && y_1 = a_0 + a_1x_{11} + a_2x_{12} + \ldots + a_kx_{1k} + \varepsilon_1 \\ &i=2 && y_2 = a_0 + a_1x_{21} + a_2x_{22} + \ldots + a_kx_{2k} + \varepsilon_2 \\ &\ldots\ldots && \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &i=n && y_n = a_0 + a_1x_{n1} + a_2x_{n2} + \ldots + a_kx_{nk} + \varepsilon_n \end{aligned} \quad (5.91.)$$

Parametrii a_j, $j=\overline{1,k}$ se numesc coeficienți de regresie și indică cu cât (mai concret cu câte unități de măsură) se modifică variabila rezultativă Y, dacă variabila factorială X_j, $j=\overline{1,k}$ se modifică cu o unitate, în condițiile în care toți ceilalți factori rămân neschimbați. Dacă un coeficient de regresie a_j este pozitiv, atunci înseamnă că variabila exogenă X_j influențează în mod direct variabila endogenă Y, iar dacă semnul coeficientului de regresie a_j este negativ, între variabila factorială X_j și cea rezultativă Y există o legătură inversă.

Sistemul (5.90) se poate scrie în formă matriceală astfel:

$$\underset{n\cdot 1}{\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}} = \underset{n\cdot(k+1)}{\begin{pmatrix} 1 & x_{11} & x_{12} & \ldots x_{1k} \\ 1 & x_{21} & x_{22} & \ldots x_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots x_{nk} \end{pmatrix}} \cdot \underset{(k+1)\cdot 1}{\begin{pmatrix} a_0 \\ a_1 \\ \ldots \\ a_k \end{pmatrix}} + \underset{n\cdot 1}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}} \quad (5.92.)$$

sau

$$Y = X\cdot A + E \quad (5.93.)$$

și

$$\hat{Y} = X\cdot A \quad (5.94.)$$

$$E = Y - \hat{Y} = Y - X\cdot A$$

Scanned with OKEN Scanner

Metoda celor mai mici pătrate presupune minimizarea expresiei $\sum_i \varepsilon_i^2$ (transcrisă matriceal ca $E' \cdot E$), dar:

$$E' \cdot E = (Y - X \cdot A)'(Y - X \cdot A) = Y' \cdot Y - 2A'(X'Y) + A \cdot X' \cdot X \cdot A \quad (5.95.)$$

$$\frac{\partial S}{A} = 0 - 2(X' \cdot Y) + 2(X' \cdot X)A = 0 \quad (5.96.)$$

$$(X' \cdot X)A = X' \cdot Y \quad (5.97.)$$

$$(X' \cdot X)^{-1}(X' \cdot X)A = (X' \cdot X)^{-1}(X' \cdot X)^{-1}(X' \cdot Y) \quad (5.98.)$$

de unde:

$$A = (X' \cdot X)^{-1} X' \cdot Y \quad (5.99.)$$

Pornind de la sistemul (5.92.), relația (5.99.) poate fi scrisă sub forma:

$$(X' \cdot X)A = X' \cdot Y \quad (5.100.)$$

unde:

X' = transpusa matricei X;

X^{-1} = inversa matricei X.

Cu relația (5.100.) se pot estima, în formă matriceală, parametrii modelului de regresie liniară multiplă.

Pentru măsurarea intensității legăturii dintre o variabilă dependentă (Y) și mai multe variabile independente, se poate folosi metoda corelației, prin determinarea raportului de corelație multiplă:

$$R_{y/x1,x2,\ldots,xk} = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}} = \sqrt{\frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}} \quad (5.101.)$$

Raportul de corelație multiplă are valori cuprinse în intervalul [0,1]. Dacă este egal cu zero, atunci nu există legătură între variabila dependentă și variabilele independente; dacă este egal cu 1, atunci există o legătură liniară perfectă între variabile. Cu cât indicatorul are valori mai apropiate de zero, cu atât legătura este mai slabă, iar cu cât are valori mai apropiate de 1, cu atât legătura este mai puternică. Raportul de corelație multiplă este mai mare decât oricare din coeficienții de corelație liniară simplă, dintre fiecare variabilă factorială și cea dependentă, luat în valoare absolută.

$$R_{y/x1,x2,\ldots,xk} > \left| r_{xj,y} \right|_{j=\overline{1,k}}, \quad (5.102.)$$

Scanned with OKEN Scanner

Pătratul raportului de corelație multiplă este *coeficientul de determinație multiplă* (R^2), care arată cât la sută din variația totală a lui Y este explicată de variabila independentă X_i ($i = \overline{1,k}$) (dacă R^2 este exprimat în procente).

Testarea semnificației raportului de corelație multiplă se poate face cu *testul F (Fisher)*, a cărui expresie este:

$$F_{calc} = \frac{n-k-1}{k} \cdot \frac{R^2}{1-R^2} \qquad (5.103.)$$

unde k este numărul variabilelor independente.

Se compară valoarea calculată a testului cu valoarea teoretică, tabelată, preluată din tabelul repartiției Fisher-Snedecor, corespunzătoare unui prag de semnificație α și gradelor de libertate k; $n-k-1$.

Dacă $F_{calc} > F_{\alpha;k;n-k-1}$ se acceptă ipoteza că variabilele independente X_1, X_2, ..., X_k au o influență semnificativă asupra variabilei rezultative Y.

Dacă se consideră influența parțială a doar uneia dintre variabilele independente X_i, asupra variabilei dependente Y, presupunându-se constantă acțiunea tuturor celorlalte variabile independente asupra lui Y, se poate calcula ***coeficientul de corelație parțială***, ce măsoară intensitatea legăturii dintre două variabile independente X_1 și X_2 asupra variabilei rezultative Y. Coeficientul de corelație parțială între Y și X_1 (considerând influența variabilei X_2 constantă) este:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{\sqrt{\left(1-r_{yx2}^2\right)\left(1-r_{x1x2}^2\right)}} \qquad (5.104.)$$

iar coeficientul de corelație parțială ce caracterizează intensitatea legăturii dintre X_2 și Y, considerând constantă influența lui X_1 este:

$$r_{yx2.x1} = \frac{r_{yx2} - r_{yx2} \cdot r_{x1x2}}{\sqrt{\left(1-r_{yx2}^2\right)\left(1-r_{x1x2}^2\right)}} \qquad (5.105.)$$

5.4. Elemente de analiză dispersională (ANOVA)

Metoda analizei dispersionale (ANOVA), cunoscută și sub denumirea de **analiză de varianță**, este utilizată pentru a verifica gradul în care valorile reale, empirice ale unei caracteristici se abat de la valorile teoretice, determinate în general cu ajutorul mediilor sau al ecuațiilor de regresie. Modelul de analiză dispersională nu explicitează relația dintre variabile, ci studiază efectul variabilei/variabilelor independente asupra celei dependente, altfel spus, măsura în care variația caracteristicii rezultative este dependentă sau nu de factorul (factorii) de grupare.

Scanned with OKEN Scanner

Analiza dispersională are la bază metoda grupării, prin ea separându-se influența factorilor esențiali (determinanți) de influența factorilor considerați întâmplători (aleatori) asupra caracteristicii „efect".

Analiza dispersională se poate efectua după un model unifactorial, bifactorial sau multifactorial, în funcție de numărul factorilor înregistrați ce-și exercită influența asupra caracteristicii rezultative (un factor, doi factori sau mai mulți factori).

În modelul de *analiză dispersională unifactorială*, care presupune gruparea valorilor în r grupe după variația factorului de grupare ales X, se testează ipoteza nulă, care promovează ideea egalității mediilor grupelor și care poate fi redată prin relația:

$$H_0 : m_1 = m_2 = ... = m_r \qquad (5.106.)$$

unde: m_i, $i = \overline{1,r}$ este media caracteristicii Y pentru grupa i, la nivelul colectivității generale, iar ipoteza alternativă, care susține inegalitatea a cel puțin două medii, poate fi redată sub forma:

$$H_1 : m_i \neq m_j \qquad (i \neq j) \qquad (5.107.)$$

Altfel spus, se testează dacă diferențele dintre mediile de grupă nu sunt prea mari pentru a fi puse doar pe seama factorilor accidentali, iar dacă există cel puțin două medii semnificativ diferite, înseamnă că factorul de grupare X are o influență semnificativă asupra variabilei rezultative Y.

Setul de date folosit în analiză este format din valorile variabilei Y, sistematizate pe cele r grupe independente, formate după variația variabilei cauzale X. Rezultatul sistematizării este evidențiat în Tabelul 5.6.

Distribuția frecvențelor variabilei *Y* în funcție de variabila *X*

Tabelul 5.6.

Grupe după X	Valorile caracteristicii rezultative Y	Numărul de valori individuale ale lui Y (volumul grupelor)	Medii de grupă condiționate de factorul de grupare
x_1	$y_{11} y_{12} \cdots y_{1j} \cdots y_1 n_1$	n_1	$\bar{y}_1$
x_2	$y_{21} y_{22} \cdots y_{2j} \cdots y_2 n_2$	n_2	$\bar{y}_2$
.....	.....	...	...
x_i	$y_{i1} y_{i2} \cdots y_{ij} \cdots y_i n_i$	n_i	$\bar{y}_i$
.....	.....	...	...
x_r	$y_{r1} y_{r2} \cdots y_{rj} \cdots y_r n_r$	n_r	$\bar{y}_r$

Pentru a stabili care ipoteză este adevărată, se iau în considerare abaterile valorilor individuale y_{ij} față de mediile grupelor (de mediile condiționate de

Scanned with OKEN Scanner

factorul de grupare) ($\bar{y}_i$), abaterile mediilor grupelor ($\bar{y}_i$) față de media generală a eşantionului ($\bar{y}$), precum şi abaterile valorilor individuale y_{ij} față de media totală a eşantionului ($\bar{y}$). Aceste abateri reflectă sintetic rezultatul modului de asociere a factorilor care determină variația caracteristicii *Y*.

Pentru identificarea ipotezei adevărate se aplică testul Fisher-Snedecor, numit pe scurt „testul F", care este calculat ca raport al unor indicatori ce reflectă variația, în funcție de sursa care produce această variație: factorul de grupare, (factorul determinant *X*, care induce variația dintre grupe) şi factorii aleatorii, neesențiali, (care induc variația din interiorul grupelor). Aşadar, statistica F se determină raportând variabilitatea dintre grupe la variabilitatea din interiorul grupelor. Pe baza repartițiilor condiționate, se determină următorii indicatori la nivelul eşantionului:

- ***mediile de grupă (condiționate)***:

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}; \qquad \text{cu } i = \overline{1,r} \tag{5.108.}$$

- ***media generală***

$$\bar{y} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{ni} y_{ij}}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} \bar{y}_i n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}}{n} \tag{5.109.}$$

unde *n* = volumul colectivității de selecție

- ***varianța dintre grupe***, numită şi ***varianță factorială*** sau ***sistematică***, este calculată ca sumă a pătratelor abaterilor mediilor de grupă de la media generală şi arată influența factorului cauzal:

$$S_1 = \sum_{i=1}^{r} (\bar{y}_i - \bar{y})^2 n_i \tag{5.110.}$$

- ***varianța din interiorul grupelor***, numită şi ***varianță reziduală***, este calculată ca sumă a pătratelor abaterilor valorilor individuale de la mediile lor de grupă şi arată influența factorilor aleatorii, reziduali asupra lui *Y*:

$$S_2 = \sum_{i=1}^{r}\sum_{j=1}^{ni} (y_{ij} - \bar{y}_i)^2 \tag{5.111.}$$

- ***varianța totală***, calculată ca sumă a pătratelor abaterilor valorilor individuale față de media totală (sau ca sumă a varianțelor anterioare), arată influența tuturor factorilor (sistematici şi aleatorii) asupra variabilei rezultative *Y*:

Scanned with OKEN Scanner

$$S = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = S_1 + S_2 \qquad (5.112.)$$

- ***dispersiile corectate***, calculate prin raportarea varianțelor la numărul de grade de libertate corespunzător:

$$s_1^2 = \frac{S_1}{r-1} = \frac{\sum_{i=1}^{r}(\bar{y}_i - \bar{y})^2 n_i}{r-1} \qquad (5.113.)$$

$$s_2^2 = \frac{S_2}{n-r} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n-r} \qquad (5.114.)$$

$$s^2 = \frac{S}{n-1} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2}{n-1} \qquad (5.115.)$$

unde *r* reprezintă numărul grupelor în care s-a împărțit colectivitatea după variația factorului sistematic *X*.

Statistica F pentru analiza dispersională unifactorială se calculează cu relația:

$$F = \frac{s_1^2}{s_2^2} \qquad (5.116.)$$

Datele se sintetizează în Tabelul 5.7.

Aplicarea modelului de analiză dispersională unifactorială în cazul distribuției de frecvențe a variabilei *Y* în funcție de variabila *X*

Tabelul 5.7.

Tipul varianței	Varianța	Nr. gradelor de libertate	Dispersia corectată	Statistica F	Gradul de influență
Factorială	$S_1 = \sum_{i=1}^{r}(\bar{y}_i - \bar{y})^2 n_i$	$r-1$	$s_1^2 = \frac{S_1}{r-1}$	$F = \frac{s_1^2}{s_2^2}$	$R^2 = \frac{S_1}{S}$
Reziduală	$S_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	$n-r$	$s_2^2 = \frac{S_2}{n-r}$		$K^2 = \frac{S_2}{S}$
Totală	$S = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$	$n-1$	$s^2 = \frac{S}{n-1}$		

Scanned with OKEN Scanner

Testul F se realizează comparând valoarea calculată anterior, a statisticii F cu valoarea tabelară, critică, F_{tab}, corespunzătoare unui nivel de semnificație α (acest lucru însemnând că rezultatele se garantează cu o probabilitate de $100(1-\alpha)\%$), și gradele de libertate $(r-1)$ și $(n-r)$. În funcție de rezultatul comparării, putem avea următoarele situții:

- dacă $F_{calc} \leq F_{tab} = F_{\alpha;\, r-1;\, n-r}$, atunci se acceptă ca adevărată ipoteza nulă H_0, deci mediile de grupă nu diferă semnificativ unele față de altele, iar eventualele diferențe ce pot apărea pot fi puse pe seama întâmplării. În acest caz, variabila rezultativă Y este independentă de factorul de grupare X și analiza dispersională reprezintă punctul final al analizei;
- dacă $F_{calc} > F_{tab} = F_{\alpha;\, r-1;\, n-r}$, atunci se respinge ipoteza nulă H_0 și se acceptă ca adevărată ipoteza alternativă, deci între mediile grupelor există o diferență semnificativă statistic, care nu este datorată numai factorilor aleatori. În acest caz variabila rezultativă depinde semnificativ de factorul de grupare X și trebuie aplicate în continuare metodele de analiză a legăturilor dintre variabile.

În cazul în care datele au fost grupate combinat, după ambele variabile, se obține o distribuție bidimensională, ce poate fi prezentată într-un tabel cu dublă intrare (Tabelul 5.8.).

Atunci formulele de calcul devin:

Sistematizarea datelor într-o distribuție bidimensională

Tabelul 5.8.

Grupe după X	Valorile caracteristicii Y	Număr total de unități	Medii condiționate ale grupelor
	$y_1 \quad y_2 \quad \dots \quad y_j \quad \dots \quad y_m$		
x_1	$n_{11} \quad n_{12} \quad \dots \quad n_{1j} \quad \dots \quad n_{1m}$	$n_{1.}$	$\bar{y}_1$
x_2	$n_{21} \quad n_{22} \quad \dots \quad n_{2j} \quad \dots \quad n_{2m}$	$n_{2.}$	$\bar{y}_2$
.....	....	.....	.....
x_i	$n_{i1} \quad n_{i2} \quad \dots \quad n_{2j} \quad \dots \quad n_{im}$	$n_{i.}$	$\bar{y}_i$
.....	....	.....	.....
x_r	$n_{r1} \quad n_{r2} \quad \dots \quad n_{rj} \quad \dots \quad n_{rm}$	$n_{r.}$	$\bar{y}_r$
Total unități	$n_{.1} \quad n_{.2} \quad \dots \quad n_{.j} \quad \dots \quad n_{.m}$	$\sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{ni} n_{.j} = n$	$\bar{y}$

$$\bar{y}_i = \frac{\sum_{j=1}^{m} y_j n_{ij}}{\sum_{j=1}^{m} n_{ij}} \qquad (5.117.)$$

Scanned with OKEN Scanner

$$\bar{y} = \frac{\sum_{j=1}^{m} y_j n_j}{\sum_{j=1}^{m} n_j} = \frac{\sum_{i=1}^{r} \bar{y}_i n_i}{\sum_{i=1}^{r} n_i} \qquad (5.118.)$$

$$S_1 = \sum_{i=1}^{r} (\bar{y}_i - \bar{y})^2 n_i \qquad (5.119.)$$

$$S_2 = \sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2 n_{ij} \qquad (5.120.)$$

$$S = \sum_{j=1}^{m} (y_j - \bar{y})^2 n_j \qquad (5.121.)$$

Exemplul 5.6. Pe baza datelor privind rezultatele generale ale Recensământului populației și locuințelor din martie 2002, s-au calculat ratele de ocupare a populației pe județe, ca raport între populația ocupată a unui județ și numărul populației în vârstă de 15 ani și peste a județului respectiv (Tabelul 5.9.).

$$r_{oc\,i} = \frac{PO_i}{P_{15i}} \cdot 100$$

Populația României pe județe și regiuni statistice, după participarea la forța de muncă în anul 2002

Tabelul 5.9. – persoane –

Regiune statistică	Județ	Populația totală	Populația activă			Populația inactivă
			Totală	din care		
				Ocupată	Șomeri	
NORD EST	Bacău	706.623	336.017	308.149	27.868	370.606
	Botoșani	452.834	213.730	189.708	24.022	239.104
	Iași	816.910	341.852	306.581	35.271	475.058
	Neamț	554.516	227.106	191.450	35.656	327.410
	Suceava	688.435	270.242	235.922	34.320	418.193
	Vaslui	455.049	187.944	163.218	24.726	267.105
Total NE		**3.674.367**	**1.576.891**	**1.395.028**	**181.863**	**2.097.476**
SUD EST	Brăila	373.174	155.810	132.804	23.006	217.364
	Buzău	496.214	177.783	140.014	37.769	318.431
	Constanța	715.151	281.907	236.068	45.839	433.244
	Galați	619.556	264.210	227.062	37.148	355.346
	Tulcea	256.492	92.471	76.589	15.882	164.021
	Vrancea	387.632	166.823	155.874	10.949	220.809

Scanned with OKEN Scanner

Total SE		**2.848.219**	**1.139.004**	**968.411**	**170.593**	**1.709.215**
SUD	Argeş	652.625	282.959	252.793	30.166	369.666
	Călăraşi	324.617	128.432	103.355	25.077	196.185
	Dâmbovița	541.763	253.604	226.764	26.840	288.159
	Giurgiu	297.859	103.444	93.666	9.778	194.415
	Ialomița	296.572	97.410	76.417	20.993	199.162
	Prahova	829.945	315.332	255.123	60.209	514.613
	Teleorman	436.025	168.102	151.091	17.011	267.923
Total S		**3.379.406**	**1.349.283**	**1.159.209**	**190.074**	**2.030.123**
SUD VEST	Dolj	734.231	311.573	281.558	30.015	422.658
	Gorj	387.308	147.931	128.440	19.491	239.377
	Mehedinți	306.732	112.616	98.879	13.737	194.116
	Olt	489.274	199.974	180.457	19.517	289.300
	Vâlcea	413.247	148.886	131.692	17.194	264.361
Total SV		**2.330.792**	**920.980**	**821.026**	**99.954**	**1.409.812**
VEST	Arad	461.791	167.024	155.728	11.296	294.767
	Caraş-Severin	333.219	121.406	104.175	17.231	211.813
	Hunedoara	485.712	196.776	163.093	33.683	288.936
	Timiş	677.926	315.004	292.507	22.497	362.922
Total V		**1.958.648**	**800.210**	**715.503**	**84.707**	**1.158.438**
NORD VEST	Bihor	600.246	232.603	215.558	17.045	367.643
	Bistrița Năsăud	311.657	157.181	142.032	15.149	154.476
	Cluj	702.755	280.174	249.281	30.893	422.581
	Maramureş	510.110	207.494	188.256	19.238	303.616
	Satu Mare	367.281	148.691	139.895	8.796	218.590
	Sălaj	248.015	96.721	84.168	12.553	151.294
Total NV		**2.740.064**	**1.122.864**	**1.019.190**	**103.674**	**1.617.200**
CENTRU	Alba	382.747	163.506	137.978	25.528	219.241
	Braşov	589.028	252.877	213.364	39.513	336.151
	Covasna	222.449	87.625	71.733	15.892	134.824
	Harghita	326.222	133.041	119.039	14.002	193.181
	Mureş	580.851	208.389	185.713	22.676	372.462
	Sibiu	421.724	164.739	145.099	19.640	256.985
Total C		**2.523.021**	**1.010.177**	**872.926**	**137.251**	**1.512.844**
BUCU-REŞTI	Municipiul Bucureşti	1.926.334	825.655	767.362	58.293	1.100.679
	SAI Ilfov	300.123	106.767	93.078	13.689	193.356
Total Bucureşti						
Total România		**21.680.974**	**8.851.831**	**7.811.733**	**1.040.098**	**12.829.143**

SURSA: Rezultatele generale ale Recensământului populației și locuințelor, martie 2002.

Scanned with OKEN Scanner

Ulterior, s-au grupat datele pe regiuni statistice, apoi a fost aplicată metoda analizei dispersionale pentru a testa dacă factorul principal de grupare (regiunea) influențează semnificativ variația ratei ocupării. Ratele medii ale ocupării, pe regiuni și pe total, calculate conform relațiilor descrise anterior sunt redate în Tabelul 5.10.

Ratele medii ale ocupării pe regiuni statistice în anul 2002

Tabelul 5.10.

Regiunea	Județe componente	Nr. de județe n_i	Rata medie a ocupării (%) $\bar{r}_{oci} = \bar{y}_i$
Nord–Est	Bacău, Botoşani, Iaşi, Neamț, Suceava, Vaslui	6	47,77
Sud–Est	Brăila, Buzău, Constanța, Galați, Tulcea, Vrancea	6	41,24
Sud	Argeş, Călăraşi, Dâmbovița, Giurgiu, Ialomița, Prahova, Teleorman	7	40,68
Sud–Vest	Dolj, Gorj, Mehedinți, Olt, Vâlcea	5	41,9
Vest	Arad, Caraş-Severin, Hunedoara, Timiş	4	42,71
Nord–Vest	Bihor, Bistrița-Năsăud, Cluj, Maramureş, Sălaj, Satu-Mare	6	46,32
Centru	Alba, Braşov, Covasna, Harghita, Mureş, Sibiu	6	42,08
Bucureşti	Municipiul Bucureşti, Sectorul agricol Ilfov	2	41,59
Total		42	43,16

Rezultatele calculelor, obținute cu produsul informatic de specialitate STATISTICA sunt prezentate în continuare.

Summary of all Effects; design: (rata_ocup.sta)

Continue... 1-REGIUNE

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	7	39,04014	34	27,31875	1,429060	,226085

Scanned with OKEN Scanner

MAIN EFFECT: REGIUNE (rata_ocup.sta)

GENERAL MANOVA Univar. Test — 1-REGIUNE

	Sum of Squares	df	Mean Square	F	p-level
Effect	273,2809	7	39,04014	1,429060	,226085
Error	928,8375	34	27,31875		

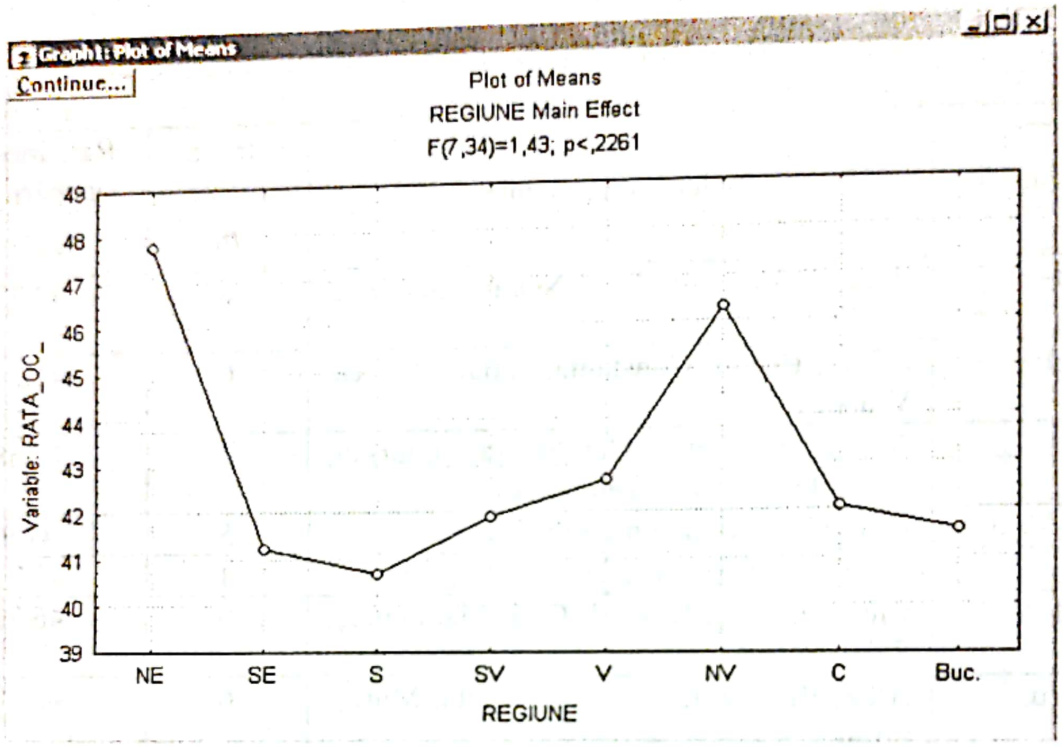


Figura 5.9. Rata ocupării pe regiuni statistice în anul 2002

Se observă că pentru un nivel de semnificație $p < 0,226085$, pentru care statistica F calculată, cu valoarea $F = 1,43$ este mai mare decât valoarea tabelată, se acceptă ca adevărată ipoteza alternativă, deci putem afirma că există diferențe semnificative între ratele de ocupare ale regiunilor statistice. Numai 22,7% din variația ratei ocupării în profil teritorial se datorează grupării pe regiuni statistice, restul de 77,3% reprezentând influența altor factori.

Scanned with OKEN Scanner

5.5. Metode neparametrice utilizate în analiza legăturilor dintre variabilele statistice

Metodele statistice prezentate până acum pentru analiza legăturilor dintre variabilele statistice sunt metode parametrice. Pentru aplicarea metodelor parametrice trebuie îndeplinite în prealabil unele condiții:

- variabilele studiate să fie numerice;
- datele să provină din populații normal distribuite după variabilele înregistrate;
- datele să provină din eşantioane de volum mare, permițând aplicarea teoremei limită centrală.

Denumirea de „metode parametrice" provine din faptul că aceste metode folosesc cei doi parametri ce definesc o distribuție de frecvențe: media şi dispersia.

Dacă însă condițiile enumerate nu sunt îndeplinite, legăturile dintre variabilele statistice nu mai pot fi studiate cu ajutorul metodelor parametrice, ci se poate recurge la metode neparametrice, mai flexibile şi mai puțin restrictive în aplicare. Astfel, metodele neparametrice se pot aplica în următoarele condiții:

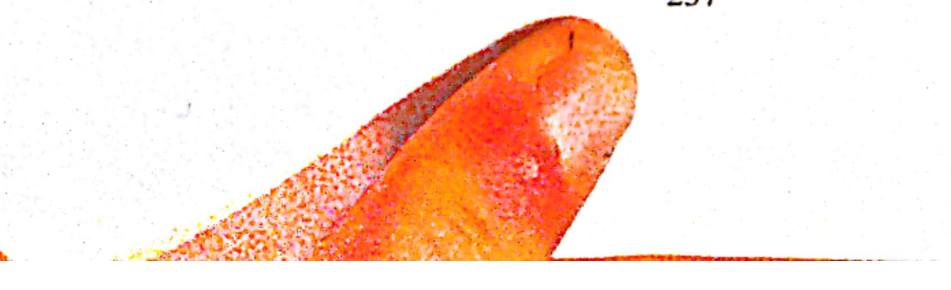
- variabilele între care se studiază legătura nu sunt numerice, ci calitative, măsurate pe scală nominală sau ordinală;
- datele nu provin din populații distribuite normal sau asimptotic normal după variabilele studiate;
- datele provin din eşantioane de volum redus, de aceea nu putem verifica normalitatea distribuției.

Deşi metodele neparametrice se folosesc în studierea legăturilor dintre variabile calitative (adică a asocierilor), ele pot fi aplicate la fel de bine şi pentru date numerice, întrucât toate metodele valabile pe o scală inferioară (calitativă) sunt valabile şi pe o scală superioară (numerică). Ele sunt, însă, mai puțin exacte, deoarece nu mai folosesc valorile în sine ale caracteristicilor, ci doar numărul de ordine (sau rangul) valorilor individuale. Unele metode neparametrice necesită, pentru a fi aplicate, doar ca datele să provină din populații continue.

Metodele neparametrice sunt uşor de aplicat, uşor de înțeles, sunt „libere de distribuție" (nu depind de forma distribuției) şi pot fi utilizate atunci când nu sunt întrunite condițiile pentru aplicarea metodelor parametrice.

5.5.1. Asocierea variabilelor nominale

Atunci când dorim să analizăm o posibilă legătură între două variabile nominale, *nealternative*, putem aplica ***testul*** χ^2.



Scanned with OKEN Scanner

Testul χ^2 a fost introdus de Karl Pearson în anii 1900 şi presupune alcătuirea unui tabel de asociere (sau de contingenţă) în care datele sunt sistematizate în r clase/grupe după o variabilă de segmentare cauzală X şi respectiv în c clase/grupe după variabila de segmentare rezultativă Y. Tabelul de contingenţă va avea, deci, r linii şi c coloane şi va fi de forma:

Tabelul de asociere (de contingenţă)

Tabelul 5.11.

Clase pt. variabila X	Clase pt. variabila Y						Total unităţi
	y_1	y_2	...	y_j	...	Y_c	
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
...	...	...	...	...	...	...	...
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r.}$
Total unităţi	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	$n_{..}$

Pentru aplicarea testului χ^2 se parcurg următorii paşi:

- formularea ipotezei nule H_0 care afirmă că nu există legătură între variabilele de segmentare (adică segmentarea populaţiei în clase după o variabilă este realizată în mod independent faţă de segmentarea populaţiei în clase după cealaltă variabilă); ipoteza alternativă va susţine, evident, contrariul.
- Alegerea nivelului de semnificaţie α şi calcularea gradelor de libertate conform relaţiei:

$$l = (r-1)(c-1) \qquad (5.122.)$$

Pe baza acestor date se preia valoarea teoretică a testului (χ^2 teoretic) din tabelul repartiţiei χ^2: $\chi^2_{\alpha;(r-1)(c-1)}$;

- Se calculează frecvenţele teoretice aşteptate, după relaţia:

$$\theta_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \qquad (5.123.)$$

unde:

$n_{i.} = \sum_{j=1}^{c} n_{ij}$ este totalul frecvenţelor de pe linia i a tabelului;

$n_{.j} = \sum_{i=1}^{r} n_{ij}$ este totalul frecvenţelor de pe coloana j a tabelului;

Scanned with OKEN Scanner

$n_{..} = n = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{c} n_{.j}$ este volumul total al colectivității.

- Se calculează testul statistic:

$$\chi^2_{calc} = \sum_{i=1}^{r}\sum_{jj=1}^{c} \frac{(n_{ij} - \theta_{ij})^2}{\theta_{ij}} \qquad (5.124.)$$

- Se compară valoarea calculată a testului cu cea teoretică; regula de decizie este următoarea:
  - dacă $\chi^2_{calc} > \chi^2_{teoretic}$, se respinge ipoteza nulă, deci considerăm că există legătură între variabilele urmărite (variabilele nu sunt independente);
  - dacă $\chi^2_{calc} \leq \chi^2_{teoretic}$, se admite ipoteza nulă, conform căreia nu există legătură între variabile (variabile independente).

În condițiile în care se stabilește că există legătură, se pune problema măsurării intensității acesteia. Pentru aceasta se calculează un coeficient de contingență (coeficientul de contingență al lui Cramer), care determină gradul de asociere între variabilele de segmentare.

$$C = \sqrt{\frac{\chi^2_{calc}}{N + \chi^2_{calc}}} \qquad (5.125.)$$

Dacă acest coeficient ia valori apropiate de zero înseamnă că între variabile există o asociere slabă (sunt aproximativ independente), iar dacă el ia valori apropiate de 1, variabilele sunt mai strâns asociate.

Exemplul 5.7. Se consideră populația ocupată a României, după forma de proprietate și după ramura de activitate, în trimestrul III 2004:

Populația ocupată pe forme de proprietate și pe activități ale economiei naționale

Tabelul 5.12.

Total	Populația ocupată în (mii persoane) după ramura de activitate:					
	Agricult. silvicult.	Industrie Construcții	Comerț turism	Transport. telecom.	Alte ramuri	
Publică	175,3	1112,5	109,1	304,0	942,5	2643,4
Privată	4731,3	1195,8	866,4	134,6	242,2	7170,3
Mixtă	81,6	639,0	64,5	49,0	37,2	871,3
Cooperatistă	2,9	38,1	12,4	0	25,5	78,9
Alta	11,0	10,2	6,2	7,9	481,1	516,4
Total	5002,1	2995,6	1058,6	495,5	1728,5	11280,3

SURSA: Ancheta asupra forței de muncă în gospodării, AMIGO, trim. III 2004, INS.

Scanned with OKEN Scanner

Întrucât ambele variabile utilizate în analiză sunt nenumerice, pentru analiza legăturii se va folosi testul χ^2, iar pentru măsurarea intensității acestei legături se va calcula coeficientul de contingență.

Frecvențele teoretice $\theta_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ sunt calculate în Tabelul 5.13.:

Tabelul 5.13.

Forma de proprietate	Frecvențe teoretice				
	Agricultură silvicultură	Industrie Construcții	Comerț turism	Transport telecomu-nicații	Alte ramuri
Publică	1172,2	702,0	248,1	116,1	405,0
Privată	3179,6	1904,1	672,9	315,0	1098,7
Mixtă	386,4	231,4	81,8	38,3	133,5
Cooperatistă	35,0	21,0	7,4	3,5	12,1
Alta	229,0	137,1	48,5	22,7	79,1

$$\chi_c^2 = \sum_{i=1}^{5}\sum_{j=1}^{5} \frac{(n_{ij} - \theta_{ij})^2}{\theta_{ij}} = 7545,1$$

$$\chi^2_{teoretic} = \chi^2_{0,05;4x4} = \chi^2_{0,05;16} = 26,3$$

Rezultă că $\chi^2_{calc} > \chi^2_{teoretic}$ (7545,1 > 26,3), deci putem concluziona că între forma de proprietate şi ramura în care îşi desfăşoară activitatea populația ocupată există o legătură, (lucru ce se poate afirma cu o probabilitate de 95%), iar intensitatea ei se poate determina cu coeficientul de contingență:

$$C = \sqrt{\frac{\chi_c^2}{N + \chi_c^2}} = \sqrt{\frac{7545,1}{11825,4}} = 0,63$$

deci între variabile există o legătură de intensitate medie.

Se observă că în trimestrul III al anului 2004 marea majoritate a populației ocupate în agricultură şi în comerț, turism îşi desfăşoară activitatea în sectorul privat, în industrie şi construcții proprietatea privată împarte supremația cu cea publică, în timp ce populația ocupată în transporturi şi telecomunicații, precum şi în alte ramuri (învățământ, sănătate, administrație publică) îşi desfăşoară activitatea preponderent în locuri de muncă din cadrul proprietății publice.

Luăm în considerare un caz particular al variabilelor nominale – şi anume variabilele nominale alternative. Studierea asocierii dintre variabile de acest tip se poate face şi cu ajutorul testului χ^2 prezentat, în care $r = 2$ şi $c = 2$.

Se obține în acest caz un tabel de asociere sau de contingență „2x2", cu următoarea formă:

Scanned with OKEN Scanner

Tabelul de asociere 2x2

Tabelul 5.14.

Clasele lui X	Clasele lui Y		Total unități
	y_1	y_2	
x_1	n_{11}	n_{12}	$n_{1.} = n_{11} + n_{12}$
x_2	n_{21}	n_{22}	$n_{2.} = n_{21} + n_{22}$
Total unități	$n_{.1} = n_{11} + n_{21}$	$n_{.2} = n_{12} + n_{22}$	$n_{..} = n_{1.} + n_{2.} = n_{.1} + n_{.2} = n_{11} + n_{12} + n_{21} + n_{22}$

După modul de distribuire a unităților statistice, după variabilele studiate, se poate determina existența și sensul asocierii:

- dacă frecvențele se concentrează pe diagonala principală a tabelului (n_{11} și n_{22}) atunci spunem că există o asociere perfectă pozitivă a variabilelor înregistrate;
- dacă frecvențele se concentrează pe diagonala secundară (n_{12} și n_{21}) atunci spunem că există o asociere perfectă negativă a celor două variabile;
- dacă produsul frecvențelor de pe diagonala principală este egal cu produsul frecvențelor de pe diagonala secundară, adică:

$$n_{11} \cdot n_{22} = n_{12} \cdot n_{21} \Leftrightarrow n_{11} \cdot n_{22} - n_{12} \cdot n_{21} = 0 \qquad (5.126.)$$

atunci spunem că variabilele sunt practic independente, între ele neexistând o asociere.

Pentru a măsura intensitatea asocierii dintre variabile, se calculează *coeficientul de asociere al lui Yule*, de forma:

$$Qa = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{n_{11} \cdot n_{22} + n_{12} \cdot n_{21}} \qquad (5.127.)$$

care ia valori în intervalul [–1,+1].

- dacă $Qa = 0$ variabilele nu sunt asociate (sunt independente);
- dacă Qa ia valori apropiate de –1 sau de +1, atunci asocierea dintre variabile este strânsă;
- dacă Qa ia valori apropiate de zero, atunci asocierea dintre variabile este slabă;
- dacă Qa este egal cu –1 sau cu +1, spunem că între variabile există o asociere perfectă;
- dacă $Qa > 0$ asocierea este pozitivă;
- dacă $Qa < 0$ asocierea este negativă.

O altă formă a coeficientului de asociere este:

$$Q_a' = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{\sqrt{n_{.1} \cdot n_{.2} \cdot n_{1.} \cdot n_{2.}}} = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{\sqrt{(n_{11} + n_{21})(n_{12} + n_{22})(n_{11} + n_{12})(n_{21} + n_{22})}} \qquad (5.128.)$$

Coeficientul determinat cu relația (5.128.) ia, de asemenea, valori în intervalul [–1,+1], cu aceeași interpretare.

Scanned with OKEN Scanner

Exemplul 5.7. Pentru datele din Tabelul 5.2., să se elaboreze tabelul de asociere în raport cu media şi să se calculeze şi interpreteze coeficientul de asociere:

Tabelul 5.15.

Clasele lui X	Clasele lui Y		Total unităţi
	$y \le 296{,}9$	$y > 296{,}9$	
$x \le 26{,}3$	$n_{11} = 5$	$n_{12} = 0$	$n_{1.} = 5$
$x > 26{,}3$	$n_{21} = 0$	$n_{22} = 5$	$n_{2.} = 5$
Total unităţi	$n_{.1} = 5$	$n_{.2} = 5$	$n_{..} = 10$

$$Qa = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{n_{11} \cdot n_{22} + n_{12} \cdot n_{21}} = \frac{25}{25} = 1$$

Între cele două variabile există, aşadar, o asociere pozitivă perfectă.

5.5.2. Asocierea variabilelor ordinale. Metoda corelaţiei rangurilor

Variabilele calitative măsurate pe o scală ordinală prezintă o anumită relaţie de ordine între variante, motiv pentru care se pot ordona şi li se pot acorda numere de ordine numite ranguri. Unitatea cea mai performantă din punctul de vedere al caracteristicii studiate primeşte rangul 1, iar unitatea cea mai puţin performantă din punctul de vedere al aceleiaşi variabile primeşte rangul n. Dacă mai multe unităţi au aceeaşi valoare sau variantă a unei caracteristici, fiecărei unităţi i se acordă un rang egal cu media aritmetică simplă a rangurilor succesive care s-ar fi acordat în mod normal. În felul acesta, nu se schimbă suma rangurilor, care este egală cu suma primelor n numere naturale $\left[\frac{n(n+1)}{2}\right]$.

Studierea asocierii între variabile ordinale se face cu ajutorul *coeficienţilor de corelaţie a rangurilor.*

Un astfel de coeficient a fost propus de C. Spearman la începutul anilor 1900, ca o alternativă la coeficientul de corelaţie liniară simplă al lui Pearson, discutat anterior. În determinarea ***coeficientului de corelaţie a rangurilor al lui Spearman***, valorile individuale ale unei variabile sunt înlocuite cu rangurile (numerele de ordine) ale variantelor. Se acordă ranguri separat pentru fiecare variabilă. Coeficientul se determină cu relaţia:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \qquad (5.129.)$$

Scanned with OKEN Scanner

unde $d_i = R_i^x - R_i^y$ reprezintă diferența dintre rangul acordat unității i după valorile lui X și cel acordat unității i după valorile lui Y;

n este numărul total de unități statistice (volumul colectivității).

Indicatorul ia valori în intervalul [–1,+1], după cum urmează:

- dacă $r_s = 0$, variabilele nu sunt asociate (sunt independente);
- dacă r_s ia valori apropiate de –1 sau de +1, atunci asocierea dintre variabile este strânsă;
- dacă r_s ia valori apropiate de zero, atunci asocierea dintre variabile este slabă;
- dacă r_s este egal cu –1 sau cu +1, spunem că între variabile există o asociere perfectă negativă sau pozitivă (caz în care va exista o identitate absolută, perfectă a rangurilor acordate după valorile lui X și după valorile lui Y: $R_i^x = R_i^y$, oricare ar fi $i = \overline{1,n}$ și deci $\sum_{i=1}^{n} d_i^2 = 0$. Dacă asocierea nu este perfectă, atunci există ranguri $R_i^x \neq R_i^y$ și deci $\sum_{i=1}^{n} d_i^2 \neq 0$.
- dacă $r_s > 0$, asocierea este pozitivă;
- dacă $r_s < 0$, asocierea este negativă.

Exemplul 5.8. Pentru datele din Tabelul 5.2., calculăm coeficientul de corelație a rangurilor Spearman. Calculele intermediare se regăsesc în Tabelul 5.16.

Tabelul 5.16.

Fabrica de încălțăminte	Număr de comenzi (x_i)	Profit (mld. lei) (y_i)	R_i^x	R_i^y	d_i^2
0	1	2	3	4	5
A	31	330	7	7	0
B	43	400	10	8	4
C	15	152	2	2	0
D	17	230	3	4	1
E	20	225	5	3	4
F	14	132	1	1	0
G	18	280	4	5	1
H	40	500	9	10	1
I	27	300	6	6	0
J	38	420	8	9	1
Total	**263**	**2969**			**12**

$$r_s = 1 - \frac{6 \cdot 12}{990} = 0{,}927$$

Scanned with OKEN Scanner

Rezultă că între variabile există o asociere puternică pozitivă (valorile mari ale numărului de comenzi se asociază cu valorile mari ale profitului, iar valorile mici ale numărului de comenzi cu valorile mici ale profitului.

Un alt indicator care se bazează pe rangurile acordate unităților statistice în funcție de valorile unei variabile ordinale este ***coeficientul de corelație a rangurilor al lui Kendall*** (r_k), care măsoară gradul în care coincid rangurile acordate după cele două variabile urmărite. Spre deosebire de coeficientul de corelație a rangurilor Spearman, cel al lui Kendall necesită ordonarea în prealabil a unităților în ordine crescătoare după rangurile acordate în funcție de valorile lui X, dar păstrând corespondența dintre rangurile perechi; altfel spus, perechile de ranguri (R_i^x, R_i^y) se reordonează crescător după R_i^x. Pentru rangurile acordate după valorile variabilei Y, se calculează:

p_i = numărul de ranguri superioare rangului curent R_i^y;

q_i = numărul de ranguri inferioare rangului curent R_i^y.

În calculul mărimilor p_i și q_i se numără doar rangurile de la poziția curentă în jos, spre capătul tabelului. Coeficientul de corelație a rangurilor Kendall se determină cu relația:

$$r_k = \frac{2S}{n(n-1)} \qquad (5.130.)$$

unde

$$S = \sum p_i - \sum q_i \qquad (5.131.)$$

Indicatorul ia valori în intervalul [–1,+1], cu aceeași interpretare ca și în cazul coeficientului lui Spearman.

În general, coeficientul lui Kendall de corelație a rangurilor are o valoare inferioară coeficientului lui Spearman și, dacă a fost luat în studiu un număr suficient de mare de unități, există relația:

$$r_k \cong \frac{2}{3} r_s \qquad (5.132.)$$

Exemplul 5.9. Pentru datele din Tabelul 5.2., calculăm coeficientul lui Kendall de corelație a rangurilor. Perechile de ranguri (R_i^x, R_i^y) determinate în Tabelul 5.16. se reordonează crescător după R_i^x, iar pentru R_i^y se calculează numărul de ranguri superioare p_i și numărul de ranguri inferioare q_i. Calculele intermediare se găsesc în Tabelul 5.17.

Scanned with OKEN Scanner

Tabelul 5.17.

Fabrica de încălțăminte	R_i^x	R_i^y	p_i	q_i
0	1	2	3	4
F	1	1	9	0
C	2	2	8	0
D	3	4	6	1
G	4	5	5	1
E	5	3	5	0
I	6	6	4	0
A	7	7	3	0
J	8	9	1	1
H	9	10	0	1
B	10	8	0	0
Total			**41**	**4**

$$r_k = \frac{2 \cdot (41 - 4)}{90} = 0{,}82$$

Valoarea indicatorului arată o asociere puternică pozitivă între variabile.

Scanned with OKEN Scanner

CAPITOLUL 6
ANALIZA EVOLUȚIEI ÎN TIMP A FENOMENELOR ECONOMICO-SOCIALE

Fenomenele din natură şi societate sunt caracterizate de dezvoltarea într-un anumit spațiu, pe de o parte, şi la un anumit moment sau interval de timp, pe de altă parte, ambele coordonate (spațiu şi timp) imprimând o serie de trăsături şi particularități fenomenelor circumscrise.

Deseori, oamenilor de afaceri, managerilor, le sunt necesare, pentru asigurarea succesului propriilor business-uri, dezvoltările (evoluțiile) în timp ale unor variabile precum profit, costuri, productivitate etc.

În prezentul capitol vom studia conceptul de „serie de timp", tipurile acestora, vom învăța să determinăm principalii indicatori statistici ce caracterizează aceste serii. Şi pentru că majoritatea fenomenelor economico-sociale sunt influențate, în dinamica lor, de o multitudine de factori, vom vedea cum îi putem identifica şi cum le putem măsura efectele. În sfârşit, pe baza dezvoltării trecute şi prezente a fenomenelor, vom putea estima (prognoza) evoluția viitoare a acestora.

6.1. Particularitățile şi tipologia seriilor cronologice

Definiție: Seria cronologică (numită şi serie de timp, sau serie dinamică) este formată dintr-un şir ordonat de valori ale unei variabile, realizate la momente sau intervale de timp succesive.

Simbolic, o serie cronologică se mai poate scrie:

$$\begin{pmatrix} 1 & 2 & \ldots\ldots\ldots & n-1 & n \\ y_1 & y_2 & \ldots\ldots\ldots & y_{n-1} & y_n \end{pmatrix}$$

unde primul şir de valori reprezintă timpul (momentele sau intervalele), iar al doilea şir reprezintă termenii seriei cronologice, constând din valorile individuale ale unei caracteristici.

Tendința generală a unei serii cronologice se poate scrie cu ajutorul unei funcții matematice, sub forma:

$$y = f(t), \qquad t = \overline{1,n} \tag{6.1.}$$

Curgerea timpului este măsurată pe scala de interval.

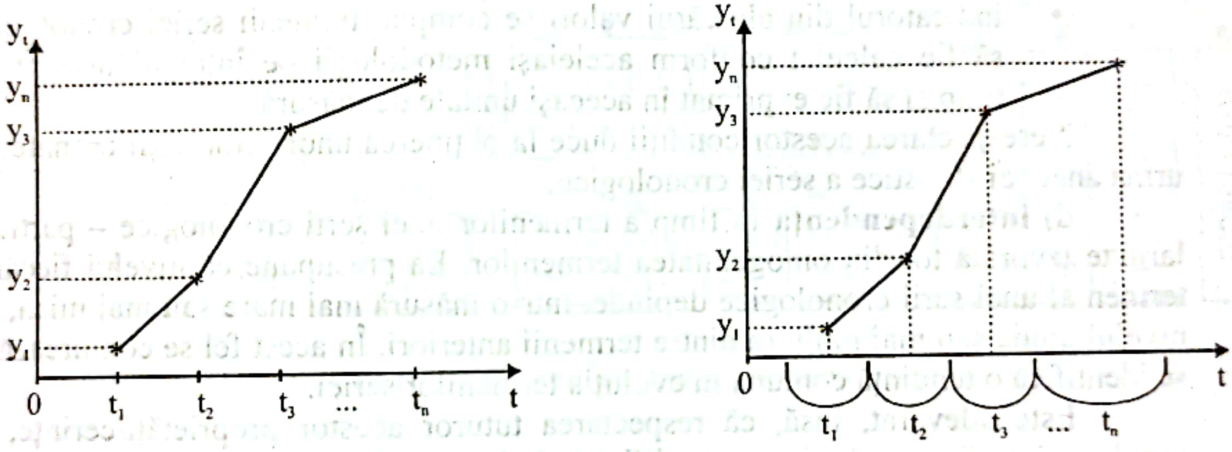
Scanned with OKEN Scanner

Dacă termenii unei serii cronologice caracterizează un interval de timp, spunem că ei sunt mărimi de flux, iar seria cronologică se numeşte **serie cronologică de intervale** (de exemplu, vânzările lunare înregistrate de o societate comercială în anul 2006).

Dacă termenii unei serii cronologice caracterizează un anumit moment de timp, atunci ei sunt mărimi de stoc, iar seria cronologică este o serie de momente (de exemplu, stocul disponibil dintr-o marfă la sfârşitul fiecărei luni).

Termenii unei serii cronologice de intervale sunt însumabili, ei constituind rezultatul unor observări statistice continue (pe zile, săptămâni, luni etc.), în timp ce termenii unei serii cronologice de momente (mărimile de stoc) nu sunt însumabili, deoarece ei pot conţine elemente (înregistrări) repetate, adică elemente ce se regăsesc la mai multe momente de timp.

Diferenţele între seriile cronologice de momente şi cele de intervale se observă şi din modul de reprezentare grafică (vezi Figura 6.1.).



a) Serie cronologică de momente

b) Serie cronologică de intervale

Figura 6.1. Tipuri de serii cronologice

Când intervalele dintre două momente succesive au lungime egală, atunci vom avea o serie cronologică de momente cu intervale egale între momente, iar atunci când intervalele dintre două momente vecine au lungime neegală avem o serie cronologică de momente, cu intervale diferite între momente.

După numărul termenilor, seriile cronologice pot fi:

- de lungime mică (predominant cu caracter de informare);
- de lungime medie;
- de lungime mare (prezintă mai mult interes din punct de vedere statistic, deoarece în cadrul unui orizont mai mare de timp este posibilă acţiunea legii numerelor mari, putând fi evidenţiată mai bine tendinţa de evoluţie a fenomenului).

Scanned with OKEN Scanner

Seria cronologică se distinge printr-o serie de particularităţi, trăsături specifice ei, între care menţionăm:

a) **variabilitatea** termenilor seriei cronologice – proprietate care reflectă modificarea valorii indicatorului (caracteristicii) la care se referă seria cronologică, de la un moment sau interval de timp la altul;

b) **omogenitatea** termenilor unei serii cronologice, cerinţă necesară în cazul calculării indicatorilor statistici ai seriei.

Această proprietate presupune că în evoluţia termenilor seriei cronologice se manifestă anumite legităţi, regularităţi, că aceşti termeni sunt de aceeaşi natură calitativă şi prezintă o dispersare minimă. Omogenitatea implică următoarea proprietate a seriilor cronologice, şi anume

c) **comparabilitatea** termenilor unei serii cronologice – trebuie să se asigure sub următoarele aspecte:

- natura calitativă a fenomenelor la care se referă seria cronologică să fie aceeaşi;
- indicatorul din ale cărui valori se compun termenii seriei cronologice să fie calculat conform aceleiaşi metodologii pe întregul orizont de timp şi să fie exprimat în aceeaşi unitate de măsură.

Nerespectarea acestor condiţii duce la obţinerea unor informaţii eronate, în urma analizei statistice a seriei cronologice;

d) **interdependenţa** în timp a termenilor unei serii cronologice – particularitate izvorâtă tot din omogenitatea termenilor. Ea presupune că nivelul fiecărui termen al unei serii cronologice depinde, într-o măsură mai mare sau mai mică, de nivelul unuia sau mai multora dintre termenii anteriori. În acest fel se conturează şi se identifică o tendinţă comună în evoluţia termenilor seriei.

Este adevărat, însă, că respectarea tuturor acestor proprietăţi/cerinţe, în cazul unei serii cronologice, este dificilă, îndeosebi în cazul seriilor cronologice mai lungi. Această contradicţie se rezolvă printr-un compromis, care să satisfacă, pe cât posibil, simultan, toate aceste proprietăţi. În analiza statistică a unei serii cronologice ne interesează: caracterizarea ei printr-un sistem de indicatori specifici, modelarea ei statistică conform particularităţilor desprinse din sistemul de indicatori şi previzionarea evoluţiei viitoare a fenomenelor.

6.2. Indicatorii seriilor cronologice

Pentru caracterizarea unei serii cronologice, se calculează, pe baza termenilor acesteia, un sistem de indicatori statistici, analitici şi sintetici care, după modul de calcul şi exprimare, pot fi structuraţi astfel:

a) indicatori absoluţi;
b) indicatori relativi;
c) indicatori medii.

Scanned with OKEN Scanner

Indicatorii pot fi primari sau derivaţi. Indicatorii derivaţi se calculează, uzual, prin comparare (sub formă de diferenţă sau sub formă de raport). Atunci când compararea se face cu primul termen al seriei (y_1) vom vorbi de indicatori cu bază fixă, iar atunci când compararea unui termen (y_t) se face cu termenul imediat anterior (y_{t-1}), vom vorbi de indicatori cu bază în lanţ (mobilă).

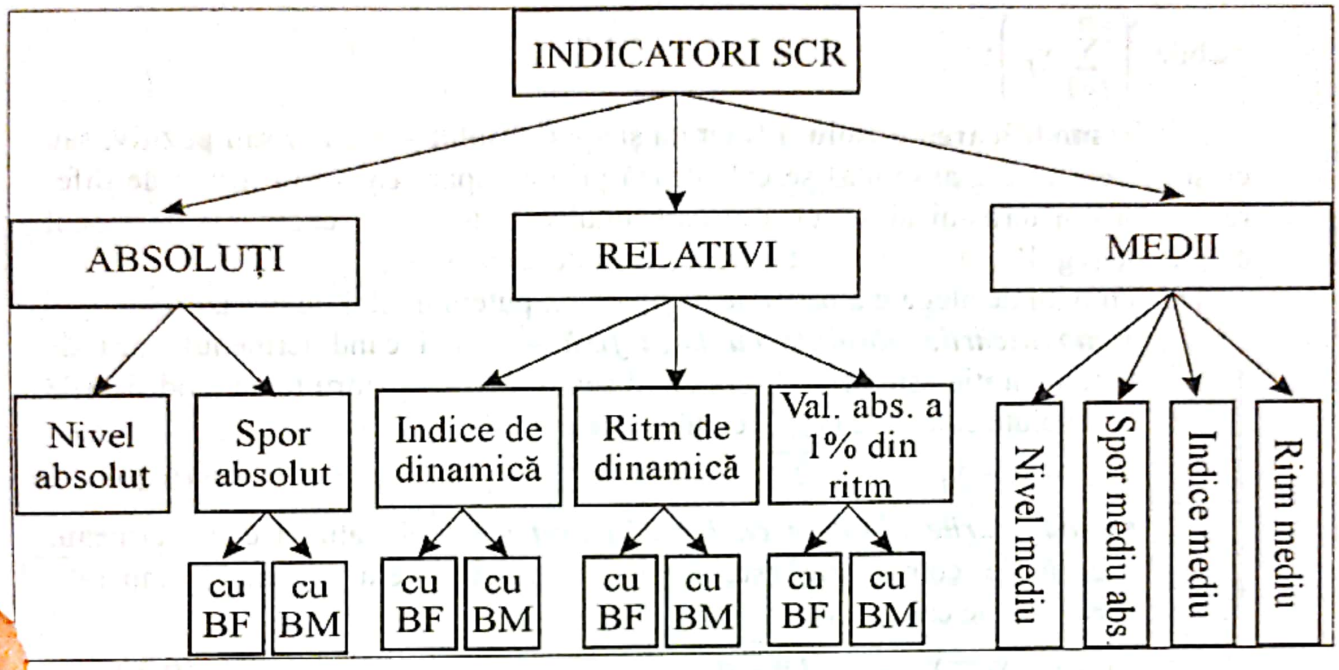


Figura 6.2. Indicatorii seriilor cronologice

unde SCR = serie cronologică;
BF = bază fixă;
BM = bază mobilă (în lanţ).

6.2.1. Indicatori absoluţi

Indicatorii absoluţi caracterizează nivelul fenomenului la care se referă seria cronologică analizată sau modificările (calculate sub formă de diferenţă), care au apărut de la un termen la altul al seriei. Aceşti indicatori se exprimă, după cum le arată şi numele, în unităţi concrete de măsură (unitatea de măsură a variabilei analizate).

Indicatorii absoluţi ai seriilor cronologice sunt:

Scanned with OKEN Scanner

a) **indicatori de nivel** – sunt reprezentați, de fapt, de termenii seriei cronologice (valorile individuale ale caracteristicii) și redau nivelul fenomenului la intervalele sau momentele de timp considerate $\{y_t\}\ t=\overline{1,n}$. Prin însumarea acestor valori individuale se poate obține un nivel totalizator, agregat al fenomenului (numai pentru seriile cronologice de intervale și numai pentru caracteristicile însumabile) $\left(\sum_{t=1}^{n} y_t\right)$;

b) **modificarea absolută** (numită și spor absolut – negativ sau pozitiv, sau creștere/descreștere absolută) se calculează prin compararea – sub formă de diferență – a doi termeni ai seriei, din care unul este termenul comparat (termenul curent, de regulă), iar celălalt – termenul bază de comparație.

După modul de alegere a bazei de comparație, putem întâlni două cazuri:

- ***modificările absolute cu bază fixă*** – atunci când termenul bază de comparație este primul termen al seriei, același pentru toate modificările absolute calculate (y_1). Relația de calcul este:

$$\Delta_{t/1} = y_t - y_1, \qquad t=\overline{2,n} \tag{6.2.}$$

- ***modificările absolute cu bază în lanț (mobilă)***, atunci când termenul bază de comparație este termenul anterior celui curent (comparat). Relația de calcul este:

$$\Delta_{t/t-1} = y_t - y_{t-1}, \qquad t=\overline{2,n} \tag{6.3.}$$

În practică, alegerea bazei de comparație nu este aleatorie, ea se face după o analiză temeinică a fenomenului și după scopul cercetării statistice. Astfel, nivelul fenomenului ales drept bază de comparație este, de regulă, unul cu o semnificație mai deosebită în dezvoltarea acestuia (de exemplu, atunci când analizăm evoluția producției pe o perioadă de timp, se alege drept bază de comparație nivelul producției înregistrat într-un an important, semnificativ pentru dezvoltarea economică – deci o bază unică fixă, sau nivelul producției atins în anul precedent fiecărui an curent din perioada analizată – bază mobilă).

Modificarea absolută arată cu câte unități s-a modificat (în sensul creșterii sau descreșterii) termenul comparat față de termenul bază de comparație.

Dacă $\Delta > 0$, atunci a avut loc o creștere, un spor.

Dacă $\Delta < 0$, atunci a avut loc o descreștere, o scădere.

Între modificările absolute cu bază fixă și cele cu bază în lanț există o serie de relații de trecere, utilizate pentru reconstituirea termenilor seriei cronologice, în situațiile în care nu se cunosc inițial toți termenii acesteia:

$$\text{a)}\ \sum_{t=2}^{k} \Delta_{t/t-1} = \Delta_{k/1}, \qquad \text{unde } k = 2, 3, \ldots, n \tag{6.4.}$$

Scanned with OKEN Scanner

Într-adevăr,

$$\sum_{t=2}^{k}\Delta_{t/t-1}=(y_2-y_1)+(y_3-y_2)+\ldots+(y_k-y_{k-1})=y_k-y_1=\Delta_{k/1}.$$

Prin însumarea modificărilor absolute cu bază în lanț a primilor k termeni se obține modificarea absolută cu bază fixă, a ultimului termen (y_k) față de primul (y_1).

b) $\Delta_{t/1}-\Delta_{t-1/1}=\Delta_{t/t-1}, \quad t=\overline{2,n}$ (6.5.)

Într-adevăr,

$$\Delta_{t/1}-\Delta_{t-1/1}=(y_t-y_1)-(y_{t-1}-y_1)=y_t-y_{t-1}=\Delta_{t/t-1}, \quad t=\overline{2,n}.$$

Prin scăderea a două modificări absolute cu bază fixă succesive se obține modificarea absolută cu bază în lanț.

Exemplul 6.1. Se consideră evoluția numărului de personal dintr-o firmă în perioada 1998-2005 (coloana 1, Tabelul 6.1.):

Tabelul 6.1.

Ani	Nr. personal	$\Delta_{t/1}$	$\Delta_{t/t-1}$	$I_{t/1}$	$I_{t/t-1}$	$R^{\%}_{t/1}$	$R^{\%}_{t/t-1}$	$A_{t/1}$	$A_{t/t-1}$
0	1	2	3	4	5	6	7	8	9
1998	50	0	–	1,0	–	0	–	0,5	–
1999	47	–3	–3	0,94	0,94	–6	–6	0,5	0,5
2000	45	v5	–2	0,9	0,96	–10	v4	0,5	0,47
2001	55	+ 5	+10	1,1	1,22	+10	+22	0,5	0,45
2002	61	+11	+ 6	1,22	1,11	+22	+11	0,5	0,55
2003	68	+18	+ 7	1,36	1,11	+36	+11	0,5	0,61
2004	70	+20	+ 2	1,4	1,03	+40	+3	0,5	0,68
2005	80	+30	+10	1,6	1,14	+60	+14	0,5	0,70

Modificările absolute ale numărului de personal au fost calculate în coloanele 2 şi 3 din Tabelul 6.1.

6.2.2. Indicatori relativi

Spre deosebire de indicatorii exprimați în mărimi absolute, în cazul indicatorilor relativi comparația se face prin **raport.**

Indicatorii relativi sunt:

a) **indicele de dinamică** (indice de creştere sau de scădere), se calculează prin raportarea termenului comparat (curent – y_t) la termenul bază de comparație. Ei se exprimă prin coeficienți sau procente şi arată de câte ori a crescut sau a scăzut nivelul fenomenului (de cât la sută) în intervalul sau momentul de timp comparat (curent) față de nivelul fenomenului în intervalul/momentul de timp ales drept bază de comparație.

Scanned with OKEN Scanner

Indicii de dinamică se pot calcula cu bază fixă şi cu bază mobilă, după relaţiile:

– ***indici de dinamică cu bază fixă***:

$$I_{t/1} = \frac{y_t}{y_1}, \quad t = \overline{2,n} \quad \text{sau} \quad I^{\%}_{t/1} = \frac{y_t}{y_1} \cdot 100 \qquad (6.6.)$$

– ***indici de dinamică cu bază mobilă***:

$$I_{t/t-1} = \frac{y_t}{y_{t-1}}, \quad t = \overline{2,n} \quad \text{sau} \quad I^{\%}_{t/t-1} = \frac{y_t}{y_{t-1}} \cdot 100 \qquad (6.7.)$$

Dacă indicele de dinamică are valori supraunitare, atunci spunem că a avut loc o creştere; dacă el are valori subunitare, s-a înregistrat o scădere a fenomenului.

Între indicii cu bază fixă şi cei cu bază mobilă există următoarele relaţii, utilizate pentru reconstituirea termenilor seriei cronologice, atunci când nu se cunosc toţi termenii acesteia:

1) produsul indicilor cu bază în lanţ, pentru primii k termeni consecutivi ai seriei este egal cu indicele de dinamică cu bază fixă, al ultimului termen (y_k) faţă de primul (y_1).

$$\prod_{t=2}^{k} I_{t/t-1} = I_{k/1} \quad , k = \overline{2,n} \qquad (6.8.)$$

Într-adevăr,

$$\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \cdot \ldots \cdot \frac{y_k}{y_{k-1}} = \frac{y_k}{y_1} = I_{k/1};$$

2) prin raportarea a doi indici de dinamică cu bază fixă, pentru două perioade succesive se obţine indicele cu baza în lanţ.

$$\frac{I_{t/1}}{I_{t-1/1}} = I_{t/t-1} \quad , t = \overline{2,n} \qquad (6.9.)$$

Într-adevăr,

$$\frac{y_t}{y_1} \cdot \frac{y_1}{y_{t-1}} = \frac{y_t}{y_{t-1}} \quad , t = \overline{2,n};$$

b) **ritmul de dinamică** (ritmul modificării relative, ritmul sporului) se determină scăzând 100% din indicele de dinamică exprimat procentual. El arată cu câte procente s-a modificat (a crescut sau a scăzut) termenul comparat faţă de termenul bază de comparaţie.

Ritmul de dinamică se poate calcula cu bază fixă sau cu bază mobilă, după relaţiile:

Scanned with OKEN Scanner

– ***Ritmul de dinamică cu bază fixă***:

$$R_{t/1} = I^{\%}_{t/1} - 100 = (I_{t/1} - 1)100 = \left(\frac{y_t}{y_1} - 1\right)100 = \frac{y_t - y_1}{y_1} \cdot 100 = \frac{\Delta_{t/1}}{y_1} \cdot 100 \qquad (6.10.)$$

– ***Ritmul de dinamică cu bază mobilă***:

$$R_{t/t-1} = I^{\%}_{t/t-1} - 100 = (I_{t/t-1} - 1)100 = \left(\frac{y_t}{y_{t-1}} - 1\right)100 = \frac{y_t - y_{t-1}}{y_{t-1}} \cdot 100 =$$

$$= \frac{\Delta_{t/t-1}}{y_{t-1}} \cdot 100 \qquad (6.11.)$$

Din relaţiile de mai sus rezultă că ritmul se mai poate calcula ca raport între modificarea absolută şi termenul bază de comparaţie (înmulţit cu 100);

c) **valoarea absolută a unui procent din ritmul de dinamică**. Deşi acest indicator se exprimă în unităţi concrete de măsură, fapt ce-l deosebeşte de ceilalţi doi indicatori relativi, el se aseamănă, totuşi, cu aceştia prin faptul că modelul de calcul cuprinde tot relaţia de raport (împărţirea). Indicatorul este egal cu raportul dintre modificarea absolută şi ritmul de dinamică, exprimat în procente.

$$A = \frac{\Delta}{R_{\%}} \qquad (6.12.)$$

Relaţia de calcul presupune repartizarea uniformă a modificării absolute pe procente de ritm, indicatorul arătând câte unităţi concrete de măsură reprezintă un procent de creştere sau de scădere.

Ca şi ceilalţi doi indicatori, şi valoarea absolută a unui procent din ritmul de dinamică se calculează cu bază fixă şi cu bază mobilă, conform relaţiilor:

– ***cu bază fixă***:

$$\left.\begin{array}{l} A_{t/1} = \dfrac{\Delta_{t/1}}{R^{\%}_{t/1}} = \\ \text{Cum} \quad R^{\%}_{t/1} = \dfrac{\Delta_{t/1}}{y_1} \cdot 100 \end{array}\right\} \Rightarrow A_{t/1} = \frac{\Delta_{t/1}}{\frac{\Delta_{t/1}}{y_1}100} = \frac{y_1}{100} = \text{constant} \qquad (6.13.)$$

Se observă deci că valoarea indicatorului cu bază fixă este constantă (deoarece baza de comparaţie e unică), indiferent de perioadă, şi egală cu a suta parte din nivelul fenomenului din primul interval/moment de timp (din nivelul-bază de comparaţie a fenomenului).

Scanned with OKEN Scanner

– *cu bază mobilă*:

$$\left.\begin{aligned} A_{t/t-1} &= \frac{\Delta_{t/t-1}}{R^{\%}_{t/t-1}} = \\ R^{\%}_{t/t-1} &= \frac{\Delta_{t/t-1}}{y_{t-1}} \cdot 100 \end{aligned}\right\} \Rightarrow A_{t/t-1} = \frac{\Delta_{t/t-1}}{\frac{\Delta_{t/t-1}}{y_{t-1}} 100} = \frac{y_{t-1}}{100} \qquad (6.14.)$$

Valoarea indicatorului cu bază mobilă nu este constantă, fiind egală cu a suta parte din nivelul fenomenului înregistrat în perioada anterioară.

Exemplul 6.2. Folosind datele din coloanele 0 şi 1 ale Tabelului 6.1., s-au calculat indicatorii relativi înscrişi în coloanele 4-9 ale aceluiaşi tabel.

6.2.3. Indicatori medii

În vederea caracterizării sintetice a unei serii cronologice, a evidenţierii tendinţei sale generale, comune, se calculează următorii indicatori medii: nivelul mediu, modificarea medie absolută, indicele mediu de dinamică, ritmul mediu de dinamică.

a) **Nivelul mediu** al seriei cronologice ($\bar{y}$) se determină în mod diferit, în funcţie de tipul acesteia (vezi Figura 6.3.).

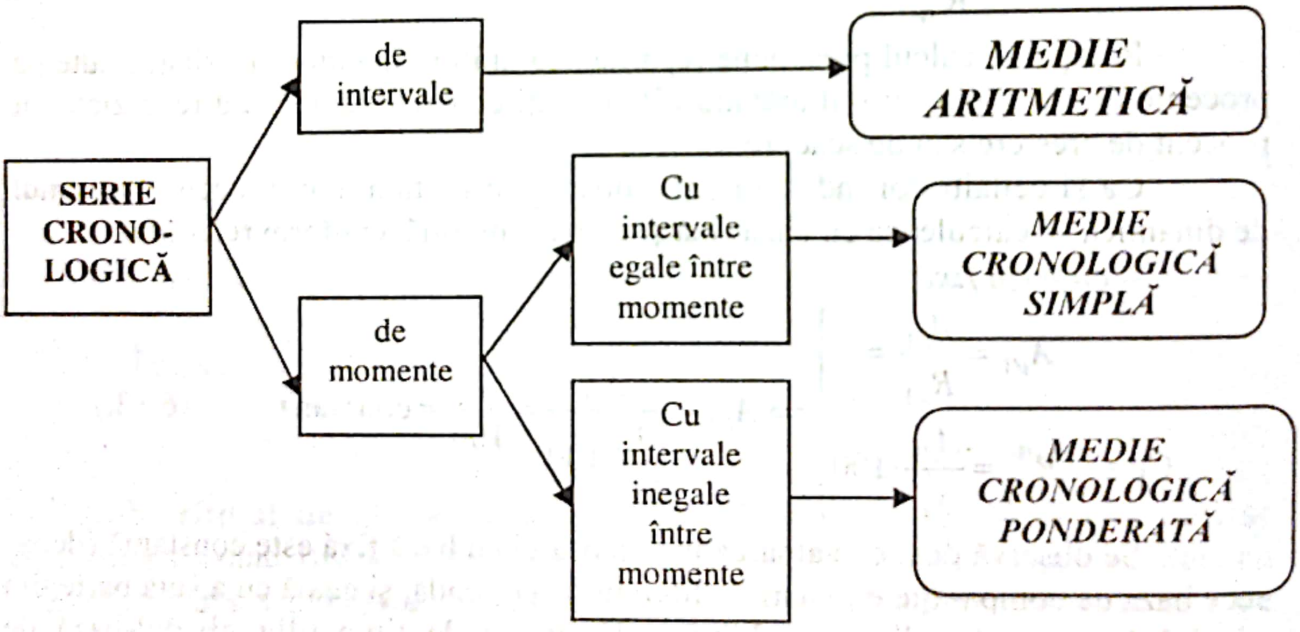


Figura 6.3. Tipurile de medii utilizate în cazul seriilor cronologice

Ca orice medie, acest indicator are sens a fi determinat numai dacă termenii din care se calculează sunt omogeni, dacă pe întreaga perioadă analizată

Scanned with OKEN Scanner

nivelul fenomenului studiat nu a prezentat variații mari. Nivelul mediu se exprimă în unități concrete de măsură.

- pentru **seria cronologică de intervale**, nivelul mediu se calculează ca medie **aritmetică simplă**, în speță prin raportul dintre suma termenilor seriei (posibilă a fi calculată, deoarece termenii seriei de intervale sunt însumabili) și numărul de termeni ai seriei cronologice.

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} \qquad (6.15.)$$

- pentru **seriile cronologice de momente**, ai căror termeni nu sunt însumabili, nivelul mediu se calculează ca **medie cronologică** (simplă sau ponderată), după cum intervalele dintre momente au lungime egală sau neegală).
  - ➢ pentru cazul seriei cronologice de momente, cu **intervale neegale între momente**, considerăm reprezentarea grafică din Figura 6.4.

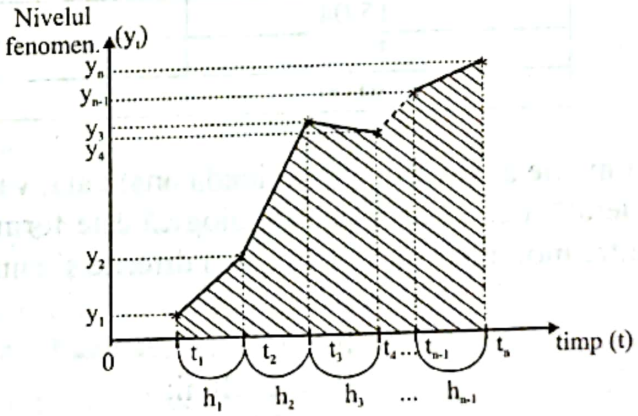


Figura 6.4. Calculul nivelului mediu al unei serii cronologice de momente

Nivelul mediu, ca medie cronologică ponderată, se calculează ca raport între aria de sub curba graficului și suma lungimii intervalelor dintre momente. Așa cum se observă, aria haşurată din Figura 6.4. se compune dintr-o sumă a ariilor $a\ (n-1)$ trapeze dreptunghice. Aria trapezului numărul i, ce are ca înălțime lungimea intervalului dintre momente t_i și t_{i+1} va fi:

$$A_i = \frac{(y_i + y_{i+1}) \cdot h_i}{2} \qquad (6.16.)$$

Scanned with OKEN Scanner

$$\bar{y}_{cr} = \frac{\sum_{i=1}^{n-1} A_i}{\sum_{i=1}^{n-1} h_i} = \frac{\frac{(y_1 + y_2)h_1}{2} + \frac{(y_2 + y_3)h_2}{2} + \frac{(y_3 + y_4)h_3}{2} + \dots + \frac{(y_{n-1} + y_n)h_{n-1}}{2}}{h_1 + h_2 + \dots + h_{n-1}} =$$

$$= \frac{y_1 \frac{h_1}{2} + y_2 \frac{h_1 + h_2}{2} + y_3 \frac{h_2 + h_3}{2} + \dots + y_{n-1} \frac{h_{n-2} + h_{n-1}}{2} + y_n \frac{h_{n-1}}{2}}{h_1 + h_2 + \dots + h_{n-1}} \qquad (6.17.)$$

unde h_i = lungimea intervalului dintre momentele t_i şi t_{i+1}, $i = \overline{1, n-1}$, exprimată în unităţi de timp.

Exemplul 6.3. Se cunosc valorile stocului dintr-un produs, din magazia unei societăţi comerciale în primele trei trimestre ale anului 2006:

Tabelul 6.2.

Momentul înregistrării stocului	Valoarea stocului (mld. lei)
1.01	10
31.03	15
15.04	7
30.06	20
30.09	25

Valoarea medie a stocului, pe perioada analizată, va fi calculată ca o medie cronologică ponderată, deoarece seria cronologică este formată din mărimi de stoc, iar intervalele dintre momente sunt de lungimi diferite şi anume:

$h_1 = 3$ luni;
$h_2 = 0{,}5$ luni;
$h_3 = 2{,}5$ luni;
$h_4 = 3$ luni.

S-a considerat că toate lunile au aceeaşi lungime (30 de zile).

$$\bar{y}_{cr} = \frac{10 \cdot \frac{3}{2} + 15 \cdot \frac{3 \cdot 5}{2} + 7 \cdot \frac{3}{2} + 20 \cdot \frac{5{,}5}{2} + 25 \cdot \frac{3}{2}}{9} = \frac{133{,}75}{9} = 14{,}86 \text{ mld.lei}$$

➢ În cazul seriei cronologice de momente, cu **intervale egale între momente**, nivelul mediu se calculează ca medie cronologică simplă, ce reprezintă un caz particular al mediei cronologice ponderate, în care $h_1 = h_2 = \dots = h_{n-1} = h$, unde h = lungimea intervalelor egale (în unităţi de timp).

Efectuându-se înlocuirea în relaţia de calcul a mediei cronologice ponderate, se obţine:

Scanned with OKEN Scanner

$$\bar{y}_{cr} = \frac{\frac{(y_1 + y_2)h}{2} + \frac{(y_2 + y_3)h}{2} + \ldots + \frac{(y_{n-1} + y_n)h}{2}}{\underbrace{h + h + \ldots + h}_{\text{de } (n-1) \text{ ori}}} =$$

$$= \frac{y_1\frac{h}{2} + y_2h + y_3h + \cdots + y_{n-1}h + y_n\frac{h}{2}}{(n-1)h} = \frac{\frac{y_1}{2} + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}}{n-1} \qquad (6.18.)$$

Exemplul 6.4. Seria cronologică prezentată în Tabelul 6.1., fiind o serie de momente, cu intervalele egale între momente, va avea media:

$$\bar{y}_{cr} = \frac{\frac{y_1}{2} + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + \frac{y_8}{2}}{7} = 58{,}7 \text{ pers./an} \approx 59 \text{ pers./an.}$$

b) **Modificarea medie absolută** ($\bar{\Delta}$) (spor mediu absolut), se determină ca medie aritmetică simplă a modificărilor absolute cu baza în lanț, cu relația:

$$\bar{\Delta} = \frac{\sum_{t=2}^{n} \Delta_{t/t-1}}{n-1} \qquad (6.19.)$$

Conform relației (6.4.), $\bar{\Delta}$ se mai scrie:

$$\bar{\Delta} = \frac{y_n - y_1}{n-1} = \frac{\Delta_{n/1}}{n-1} \qquad (6.20.)$$

Indicatorul arată cu câte unități concrete de măsură s-a modificat nivelul fenomenului, în medie, pe o unitate de timp din orizontul analizat.

Deoarece modificarea medie absolută este o mărime medie, ea are conținut real doar dacă termenii din care s-a calculat sunt omogeni, altfel spus, dacă **modificările absolute cu baza în lanț au valori apropiate unele față de altele.** Aceasta cu atât mai mult cu cât se observă că indicatorul $\bar{\Delta}$ se poate determina [conform relației (6.20.)] doar pe baza primului și ultimului termen al seriei, fără luarea în considerare și a termenilor intermediari. În plus, dacă acești termeni extremi prezintă valori mult diferite față de termenii intermediari, nici $\bar{\Delta}$ nu mai prezintă conținut real. Această condiție limitează aria de aplicabilitate a indicatorului.

c) **Indicele mediu de dinamică** ($\bar{I}$) (de creștere sau de scădere) se determină ca medie geometrică a indicilor cu bază în lanț, adică:

$$\bar{I} = \sqrt[n-1]{\prod_{t=2}^{n} I_{t/t-1}} \qquad (6.21.)$$

Scanned with OKEN Scanner

Conform relației (6.5.) $\bar{I}$ se mai poate scrie:

$$\bar{I} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{I_{n/1}} \qquad (6.22.)$$

Indicatorul arată, printr-o valoare sintetică, de câte ori s-a modificat, în medie, nivelul fenomenului analizat pe orizontul de timp luat în calcul.

Observațiile privind reprezentativitatea mediei făcute la modificarea medie absolută sunt valabile și aici; pentru ca indicele mediu de dinamică să aibă conținut real (să fie reprezentativ), este necesar ca toți termenii din care se calculează să fie omogeni, adică **indicii de dinamică cu bază în lanț să aibă valori apropiate**, cerință cu atât mai necesară, cu cât se observă că indicatorul se poate calcula doar cu ajutorul termenilor extremi ai seriei cronologice.

d) **Ritmul de dinamică** ($\bar{R}$) (modificare medie relativă, sau spor mediu relativ, sau rată medie de creștere sau scădere) se calculează scăzând 100% din indicele mediu de dinamică exprimat procentual:

$$\bar{R} = \bar{I}_{\%} - 100 = (\bar{I} - 1) \cdot 100 \qquad (6.23.)$$

Indicatorul arată cu câte procente crește sau scade, în medie, nivelul fenomenului analizat, de la o perioadă la alta, pe întregul orizont de timp.

Exemplul 6.5. Pentru seria cronologică din Tabelul 6.1., s-au calculat următorii indicatori medii:

$$\bar{\Delta} = \frac{y_{99} - y_{92}}{7} = \frac{80 - 50}{7} = \frac{30}{7} = 4{,}3 \approx 4 \text{ persoane/an}$$

$$\bar{I} = \sqrt[7]{\frac{80}{50}} = 1{,}07 \quad (107\%)$$

$$\bar{R} = +7\%$$

În medie, firma are într-un an 59 de persoane. Numărul de personal a crescut în perioada analizată cu 4 persoane în medie pe an, adică de 1,07 ori, ceea ce înseamnă o creștere cu 7 % în medie, anual.

6.3. Componentele unei serii cronologice

Manifestarea în timp a fenomenelor economico-sociale se realizează sub incidența a numeroși factori, care ne pot ajuta să explicăm comportamentul variabilelor economice pe un anumit orizont temporal.

Am văzut că una din cele mai importante proprietăți ale seriilor cronologice este variabilitatea, care poate conferi fenomenului o mișcare, o deplasare

Scanned with OKEN Scanner

continuă, unidirecţională, la care se adaugă şi unele oscilaţii aparent întâmplătoare, pe termen mai scurt sau mai lung.

Analizând componentele seriilor cronologice, putem explica evoluţia trecută şi prezentă a fenomenelor economico-sociale, dar – scop mult mai important – avem posibilitatea să previzionăm şi comportamentul fenomenului într-un viitor mai mult sau mai puţin depărtat, prin proiectarea „în viitor" a modelelor de evoluţie şi comportament identificate în trecut. Aceasta implică o presupunere fundamentală: aceea că aceste „modele", „tendinţe", care s-au manifestat în perioada trecută vor fi valabile, relevante şi în viitor. Desigur că această presupunere (ipoteză) este mai aproape de adevăr pentru perioada imediat următoare, însă, pe un orizont mai mare de timp – ar putea interveni şi alţi factori, care ar putea modifica comportamentul prestabilit.

În ciuda naturii aparent întâmplătoare a manifestării variabilelor într-un interval de timp, deseori există şi se poate identifica un model de bază, teoretic, fundamental, ce cuprinde diferite elemente, reflectă influenţa anumitor factori, astfel încât el să reproducă cât mai fidel seria de timp originală, empirică.

Termenii unei serii cronologice se descompun, de obicei, în următoarele componente:

a) **Trendul** (T) – reprezintă tendinţa generală ($\hat{y}$), ce corespunde unei evoluţii, mişcări generale sistematice, fundamentale, sesizabile pe perioade lungi de timp, generate de acţiunea unor factori cu acţiune de lungă durată.

Trendul constituie componenta principală a termenilor seriei cronologice.

b) **Oscilaţiile periodice sezoniere** (S) reprezintă fluctuaţii regulate, care se repetă în cadrul unei perioade complete pe până la un an.

Astfel, dacă avem de analizat date trimestriale, atunci fluctuaţiile vor fi asociate fiecărui trimestru, iar dacă datele ce trebuie studiate sunt zilnice – fluctuaţiile vor fi asociate fiecărei zile etc. De exemplu, cererea trimestrială de jocuri electronice, pe 5 ani, este o serie cronologică afectată de factorul sezonier, deoarece ea atinge un nivel maxim în trimestrul IV al fiecărui an – din pricina sărbătorilor de iarnă – şi un nivel minim în trimestrul III – ce corespunde vacanţelor, când familiile sunt, de regulă, plecate de acasă.

Oscilaţiile sezoniere se repetă, cu o periodicitate constantă de până la un an şi ele apar atunci când seriile cronologice se referă la perioade de timp mai mici decât anul (trimestrul, luna, săptămâna etc.).

Factorii sezonieri care produc oscilaţiile sezoniere pot fi:

- factori naturali, climatici (de exemplu, afectate de asemenea factori pot fi: producţia agricolă, vânzările de băuturi răcoritoare, vânzările de articolele de îmbrăcăminte etc.);

Scanned with OKEN Scanner

- factori sociali (de exemplu, tradiții, sărbători, concedii). Aceştia pot afecta: activitatea turistică, vânzările de rechizite şcolare, de pomi de iarnă etc.

Pot exista şi fluctuații sezoniere care pot fi sesizate pe perioade de timp mai mici decât trimestrul (săptămâna, ziua). De exemplu, consumul orar de energie electrică sau de apă menajeră în cadrul unei zile.

c) **Oscilațiile periodice ciclice** (C) reprezintă fluctuații regulate, pe termen mai lung, care pot deveni complete în decursul câtorva ani. Cauzele care produc aceste oscilații pot fi naturale (oscilațiile producției agricole, sub impactul ciclurilor meteorologice), sau de natură economico-socială (ciclurile economice sau ciclurile de afaceri – provocate de modernizarea şi înlocuirea aparatului de producție, de reproducerea materiilor prime, a resurselor naturale, inovațiile, succesele în cercetarea științifică, războaiele).

d) **Abateri aleatorii (sau reziduale)** (ε), sunt oscilații accidentale față de linia de trend, ce apar sub influența unor factori imprevizibili (greve, conflicte de muncă apărute spontan şi neprevăzut, calamități naturale – inundații, cutremure, războaie etc.). Tot din categoria factorilor aleatorii fac parte şi erorile de observare ce apar în procesul de cercetare statistică.

În cercetarea statistică a unei serii cronologice nu se delimitează (identifică), de cele mai multe ori, componenta ciclică, ci atenția se îndreaptă spre punerea în evidență a celorlalte trei componente: trendul, componenta sezonieră şi componenta aleatorie. De asemenea, nu toate seriile cronologice prezintă neapărat toate componentele menționate anterior. Astfel, termenilor unei serii cronologice le poate lipsi trendul, caz în care seria se mai numeşte „staționară", prezentând doar oscilații aleatorii şi sezoniere. La o astfel de serie, termenii oscilează în jurul unei valori constante fixe, egale cu media (Figura 6.5.).

Pentru a reconstitui termenii seriei cronologice, cele patru componente se pot combina după două modele: aditiv şi multiplicativ.

1) **Modelul aditiv** de combinare a componentelor unei serii cronologice

Modelul aditiv constă în însumarea celor patru componente, pentru a regăsi termenii reali ai seriei cronologice (y_t).

$$y_t = \hat{y}_t + S_t + C_t + \varepsilon_t \quad t = \text{timpul} \qquad (6.24.)$$

Deoarece în majoritatea cazurilor nu se dispune de date suficiente pentru identificarea elementului ciclic, modelul va fi redus la forma:

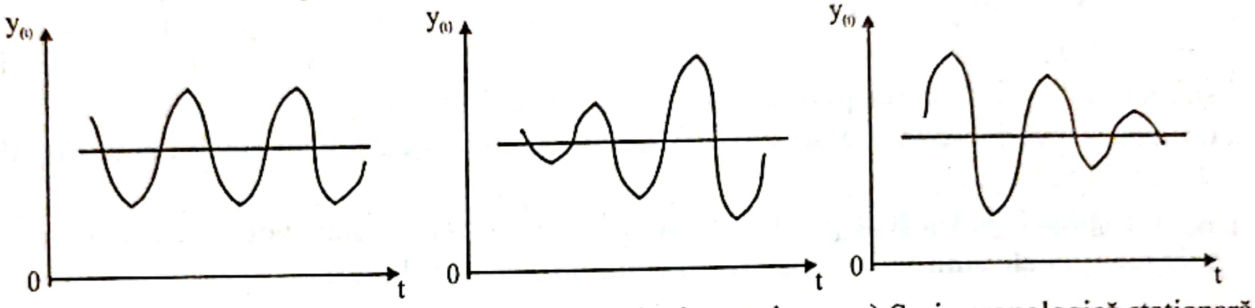
$$y_t = \hat{y}_t + S_t + \varepsilon_t \qquad (6.25.)$$

Întrucât oscilațiile sezoniere se repetă identic în fiecare subperioadă j din cadrul perioadei i (de exemplu, perioada poate fi anul, iar subperioada – trimestrul), modelul devine:

Scanned with OKEN Scanner

$$y_{ij} = \hat{y}_{ij} + S_j + \varepsilon_{ij} \quad i = \overline{1,m} \quad j = \overline{1,p} \qquad (6.26.)$$

Deoarece elementul aleatoriu este imprevizibil, se presupune că abaterile pe care le determină factorul aleatoriu, se compensează reciproc şi că media componentei aleatorii este zero. Utilizarea modelului aditiv este indicată atunci când amplitudinea oscilaţiilor faţă de linia de trend este aproximativ constantă pentru aceeaşi subperioadă (trimestru, de exemplu) a fiecărei perioade complete (a fiecărui an, de exemplu) (Figura 6.6.).



a) Serie cronologică staţionară cu oscilaţii de amplitudine constantă

b) Serie cronologică staţionară cu oscilaţii de amplitudine crescătoare

c) Serie cronologică staţionară cu oscilaţii de amplitudine descrescătoare

Figura 6.5. Tipuri de serii cronologice staţionare

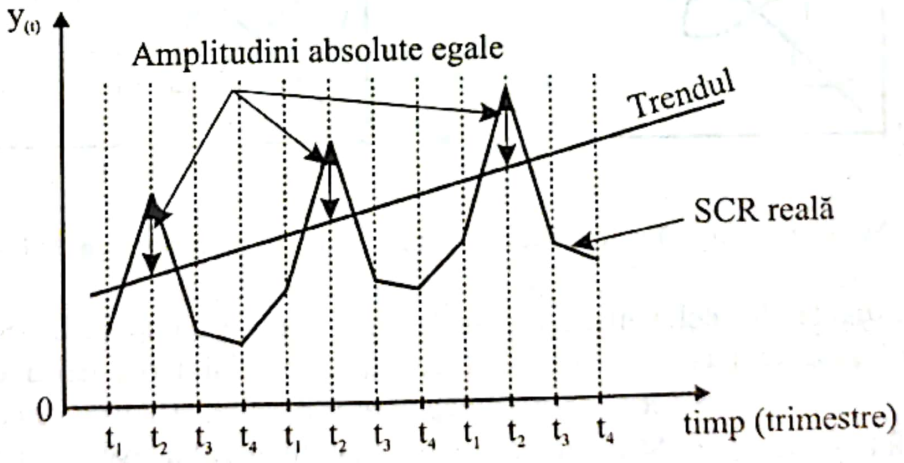


Figura 6.6. Serie cronologică cu oscilaţii egale faţă de trend

2) **Modelul multiplicativ** de combinare a componentelor unei serii cronologice

Prin modelul multiplicativ, obţinerea termenilor reali ai seriei cronologice se face prin înmulţirea celor 4 componente menţionate anterior. Cum componenta ciclică nu se poate, în general determina, modelul se scrie:

Scanned with OKEN Scanner

$$y_{ij} = \hat{y}_{ij} * S'_j * \varepsilon'_{ij} \quad i = \overline{1,m} \qquad (6.27.)$$
$$j = \overline{1,p}$$

În acest model, doar valoarea empirică a termenilor seriei cronologice (y_{ij}) şi cea a trendului ($\hat{y}_{ij}$) au valori absolute, concrete, (şi totodată au aceleaşi valori ca şi în modelul aditiv), pe când celelalte componente – sezonieră şi reziduală (S'_j şi ε'_{ij}) sunt rezultatul unor rapoarte şi au valori diferite faţă de modelul aditiv.

În cadrul modelului multiplicativ, componenta sezonieră are valoarea neutră *1*, iar insuficienţa datelor face imposibilă, uneori, identificarea componentei ciclice.

Modelul multiplicativ este indicat a fi utilizat atunci când amplitudinea oscilaţiilor faţă de trend este egală cu un acelaşi procent din valoarea trendului, în aceeaşi subperioadă *j* a fiecărei perioade *i*. Mai general, putem spune că acest model se utilizează când oscilaţiile se amplifică sau se micşorează faţă de linia de trend [Figura 6.5. b) şi c)].

În Figura 6.7. s-a luat pentru exemplificare un trend liniar, pentru claritate, dar cele menţionate sunt valabile şi în cazul unui trend non-liniar.

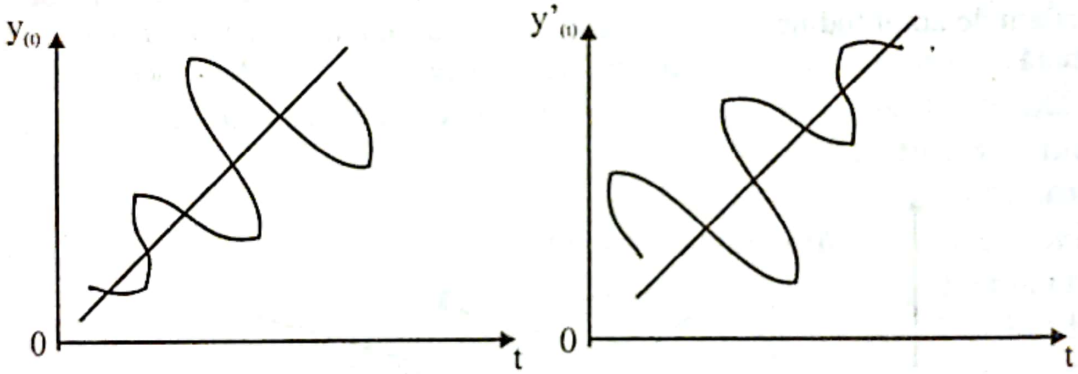


Figura 6.7. Serie cronologică cu oscilaţii: a) amplificate; b) atenuate faţă de linia de trend

Alegerea modelului celui mai adecvat de combinare a componentelor termenilor seriei cronologice se face după ce s-a efectuat în prealabil o analiză complexă a fenomenului în cauză şi a seriei cronologice care îl descrie. Analiza va cuprinde şi o reprezentare grafică, ce poate facilita alegerea modelului celui mai potrivit. În majoritatea cazurilor nu există un model perfect care să reconstituie exact seria cronologică, dar se poate alege modelul care aproximează mai bine seria de valori reale.

Scanned with OKEN Scanner

6.4. Metode de determinare a tendinței de lungă durată

Analiza unei serii cronologice debutează cu determinarea trendului (a tendinței de lungă durată, sau a tendinței centrale) a seriei. Trendul este componenta de bază a unei serii cronologice, ce apare ca urmare a acțiunii unor factori esențiali, de lungă durată.

Definiție: Identificarea trendului ca o serie de valori și înlocuirea termenilor reali ai seriei cu valorile trendului obținute prin diferite modele se numește ajustare (sau netezire a seriei cronologice).

Denumirea de „netezire" vine de la faptul că prin operația de ajustare se „aplatizează" evoluția fenomenului, prin eliminarea oscilațiilor sezoniere și reziduale.

Trendul este expresia manifestării legităților care guvernează dezvoltarea fenomenului; pe baza trendului se poate identifica și prognoza evoluția viitoare a SCR.

Metodele de determinare a trendului SCR se împart în două mari categorii: metode mecanice și metode analitice.

6.4.1. Metode mecanice de determinare a trendului unei serii cronologice

Metodele elementare de determinare a trendului unei serii cronologice sunt:

a) metoda grafică;
b) metoda mediilor mobile;
c) metoda modificării medii absolute;
d) metoda indicelui mediu de dinamică.

a) Metoda grafică

Analiza prealabilă a oricărei serii cronologice poate începe cu reprezentarea sa grafică, deoarece aceasta oferă o imagine sugestivă a evoluției în timp a fenomenului oglindit de seria cronologică. Modalitățile de reprezentare grafică ale seriilor cronologice sunt numeroase și variate, dintre ele menționăm:

I. **Cronograma (historiograma)** – este, așa cum îi arată și numele, reprezentarea grafică tipică, specifică a unei serii cronologice. Ea se trasează într-un sistem de axe rectangulare, de obicei în cadranul întâi al acestuia. Pe cele două axe se vor reprezenta: timpul – pe abscisă (se marchează momentele sau intervalele – după cum seria este formată din mărimi de stoc sau de flux), iar termenii seriei – pe ordonată (Figura 6.8.).

Scanned with OKEN Scanner

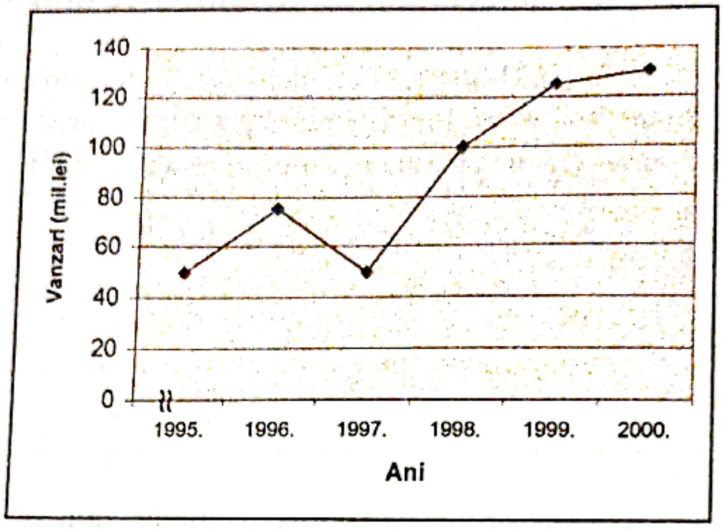


Figura 6.8. Cronograma (historiograma)

Termenii seriei cronologice se figurează ca puncte în plan, care se unesc, de regulă, prin segmente de dreaptă. Alura graficului depinde în mare măsură de alegerea scărilor de reprezentare, a unităților de lungime ale celor două scări, de stabilirea justă a raportului dintre ele.

Atât în cazul cronogramei, cât şi în cazul celorlalte tipuri de reprezentări grafice ale seriilor cronologice, se pot utiliza două tipuri de scări:

- **scara uniformă (aritmetică)** – caracterizată prin faptul că diferențele egale între termenii seriei au drept corespondent segmente de dreaptă pe axa ordonatelor şi se recomandă a fi utilizată atunci când termenii seriei alcătuiesc aproximativ, o progresie aritmetică;
- **scara semilogaritmică** – caracterizată prin aceea că pe abscisă – pentru reprezentarea timpului se foloseşte o scară aritmetică, iar pe ordonată – pentru reprezentarea nivelului termenilor seriei se utilizează o scară logaritmică. În cadrul scării logaritmice, distanțele sunt proporționale cu logaritmii termenilor reprezentați.

Acest tip de scară se recomandă a fi folosită în următoarele cazuri:

- când termenii seriei cronologice se dispun aproximativ într-o progresie geometrică;
- când variația termenilor seriei prezintă valori mari şi foarte mari de-a lungul orizontului de timp analizat;
- când se compară, cu ajutorul graficului, evoluțiile unor indicatori economico-sociali ce prezintă diferențe mari de mărime între ei.

Scanned with OKEN Scanner

Dacă termenii seriei cronologice nu încep de la zero (originea axelor), ci de la o valoare sensibil mai mare, se face o întrerupere a scării şi a graficului, pe axa ordonatelor, prima gradaţie pe ordonată, deasupra întreruperii, fiind o valoare rotundă mai mică, dar apropiată de valoarea minimă a termenilor seriei.

Dacă avem de analizat un sistem de serii cronologice, acestea pot fi reprezentate în acelaşi grafic. Dacă termenii tuturor seriilor cronologice se exprimă prin aceeaşi unitate de măsură, atunci se poate alege o singură scară de reprezentare pe axa ordonatelor. Dacă însă termenii seriilor cronologice sunt exprimaţi în unităţi de măsură diferite, atunci se vor alege mai multe scări de reprezentare diferite, câte una pentru fiecare serie.

În acest ultim caz va fi, de asemenea, necesară o legendă, care să explice semnificaţia simbolurilor folosite în trasarea curbelor. Pentru a elimina problemele ce apar atunci când termenii seriilor cronologice sunt exprimaţi în unităţi de măsură diferite, se pot înlocui termenii absoluţi cu termenii relativi ai aceleiaşi serii. Primul termen al noii serii ce va reprezenta baza, va fi egal cu 1 sau 100, iar ceilalţi termeni se recalculează, prin împărţire la termenul-bază.

II. **Diagrama prin coloane** – în care timpul se reprezintă pe abscisă, iar termenii seriei pe ordonată. În cazul seriilor cronologice de momente, din fiecare moment se ridică câte o coloană, a cărei înălţime este proporţională cu nivelul termenilor. În cazul seriilor cronologice de intervale, coloanele vor fi plasate pe mijlocul segmentelor de pe abscisă care corespund intervalelor. De obicei este indicat ca toate coloanele să aibă aceeaşi lăţime, iar distanţele dintre ele să fie proporţionale cu distanţele dintre momente. În loc de coloane se pot folosi alte desene simbolice, pentru a reprezenta termenii seriei (piramide, conuri etc.) (Figura 6.9).

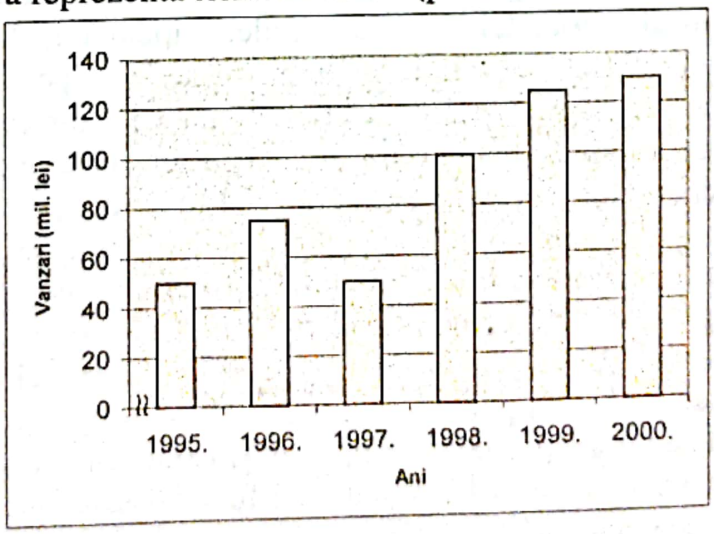


Figura 6.9. Diagrama prin coloane

Scanned with OKEN Scanner

III. **Diagrama prin benzi** este recomandată a se folosi atunci când se reprezintă (simultan) termenii unor serii cronologice, termeni care constituie nişte indicatori strâns legaţi între ei (de exemplu, venituri-cheltuieli, profit-pierderi, import-export etc.). Benzile au drept bază axa timpului (axa ordonatelor), iar lungimea lor, pe orizontală, va fi proporţională cu mărimea termenilor seriei (Figuri 6.10.).

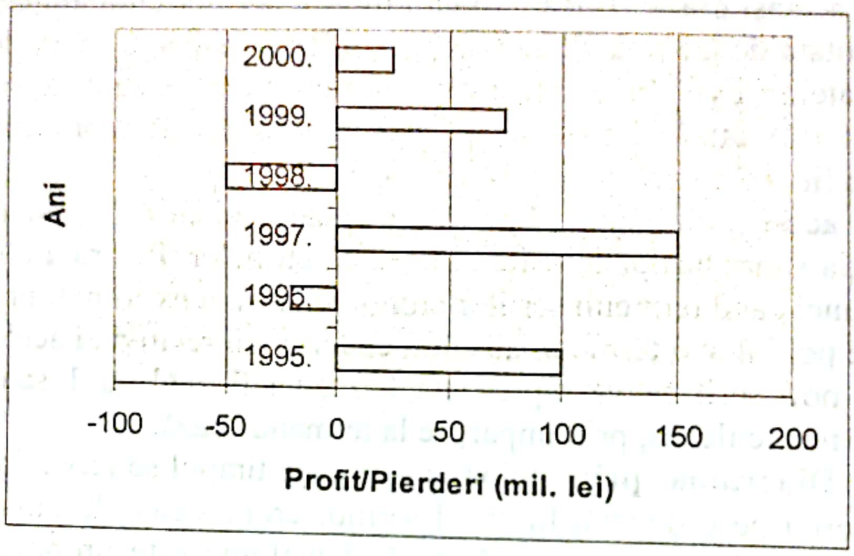


Figura 6.10. Diagrama prin benzi

IV. **Diagrame polare** (numite şi **diagrame radiale** sau **diagrame în spirală**) se construiesc cu ajutorul reţelelor radiale şi se utilizează în special în reprezentarea seriilor cronologice afectate de fluctuaţii sezoniere. Astfel, ciclurile periodice din cadrul perioadei analizate (zile, săptămâni, luni, trimestre) se marchează pe circumferinţa cercului, iar termenii seriei – pe razele cercului. Un exemplu de diagramă polară s-a reprezentat în Figura 2.19. din capitolul 2.

Graficul ocupă un loc important în analiza seriilor cronologice, el servind la popularizarea datelor şi informaţiilor (în special în rândul persoanelor neavizate, fără pregătire de specialitate), la identificarea modalităţilor de prelucrare, la alegerea funcţiei analitice de modelare a trendului seriilor cronologice.

Metoda grafică de determinare a trendului este o metodă mecanică de ajustare, prin care se realizează o ajustare vizuală. Aplicarea ei constă în unirea – printr-o dreaptă sau curbă – a punctelor extreme ale seriei cronologice, astfel încât abaterile, diferenţele între valorile de pe dreaptă şi valorile reale să fie minime.

Ajustarea vizuală stă la baza stabilirii funcţiei sau funcţiilor care aproximează optim tendinţa generală a seriei.

Scanned with OKEN Scanner

b) **Metoda mediilor mobile (metoda mediilor glisante, alunecătoare)**

Constă în înlocuirea termenilor reali ai seriei cronologice cu valori teoretice, numite **medii mobile**. Metoda se bazează pe proprietatea mediei aritmetice de compensare a erorilor, deci ea înlătură influenţa factorilor ce determină oscilaţiile periodice. Şirul de medii mobile obţinut reflectă tendinţa comună, obiectivă, generală a seriei.

Mediile mobile se calculează ca medii aritmetice parţiale dintr-un anumit număr de termeni succesivi ai seriei. Acest număr (p) depinde de periodicitatea oscilaţiilor (Figura 6.11.).

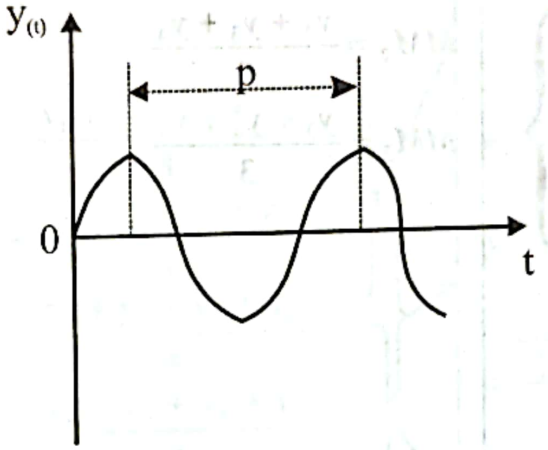


Figura 6.11. Determinarea numărului de termeni ai seriei cronologice din care se calculează mediile mobile

După ce s-a calculat o medie mobilă dintr-un set de termeni succesivi, următoarea medie mobilă se va obţine dintr-un nou set de termeni, provenit din vechiul set prin deplasarea în timp cu un termen (adică, altfel spus, noul set de termeni din care se calculează următoarea medie mobilă este format din vechiul set, din care se elimină primul termen şi se include termenul următor vechiului set).

- Dacă media mobilă se calculează dintr-un număr impar (de exemplu, $p = 3$) de termeni, schema de calcul a mediilor mobile este următoarea (Tabelul 6.3.):

Se observă că mediile mobile astfel calculate sunt de fapt termenii ajustaţi ai seriei şi că ele se plasează în dreptul unor termeni reali (MM_1 în dreptul lui y_2, MM_2 în dreptul lui y_3 etc.). Numărul de medii mobile obţinut este mai mic decât numărul de termeni reali ai seriei (în cazul nostru sunt cu două mai puţine: MM_1, MM_2, ..., MM_{n2}). Primul şi ultimul termen real nu vor avea corespondent o valoare ajustată, adică o medie mobilă. Cu cât numărul de termeni din care se calculează media mobilă este mai mare, cu atât se vor obţine mai puţine medii mobile şi se pierde mai multă informaţie. În acest caz, netezirea va fi mai drastică, mai accentuată, mai puternică.

Scanned with OKEN Scanner

Calculul mediilor mobile dintr-un număr impar de termeni

Tabelul 6.3.

Termenii reali ai seriei cronologice	Medii mobile (MM_i) $i=\overline{1,n-2}$	Valori ajustate
y_1		
y_2	$MM_1 = \frac{y_1 + y_2 + y_3}{3}$	$\hat{y}_1 = MM_1$
y_3	$MM_2 = \frac{y_2 + y_3 + y_4}{3}$	$\hat{y}_2 = MM_2$
y_4	$MM_3 = \frac{y_3 + y_4 + y_5}{3}$	$\hat{y}_3 = MM_3$
y_5		
⋮		
y_{n-2}	$MM_{n-2} = \frac{y_{n-2} + y_{n-1} + y_n}{3}$	$\hat{y}_{n-2} = MM_{n-2}$
y_{n-1}		
y_n		

Pentru cazul general, prin această metodă se vor obține $k = n - (p-1)$ medii mobile,
unde

n = numărul de termeni ai seriei cronologice;
p = numărul de termeni din care se calculează media mobilă.

Prin acest procedeu se pierd $\frac{p+1}{2}$ termeni reali ai seriei. Tendința generală de evoluție a seriei cronologice ajustate prin această metodă este exprimată de șirul mediilor mobile calculate.

- Dacă mediile mobile se calculează dintr-un număr par de termeni, mediile mobile determinate nu se vor plasa în dreptul unor termeni reali ai seriei, deci nu-i pot înlocui, de aceea se impune o operație suplimentară, de **centrare** a mediilor mobile. Pentru aceasta se determină noi

Scanned with OKEN Scanner

medii aritmetice, calculate din câte două medii mobile consecutive, care se vor plasa în dreptul unor termeni reali, pe care- i vor înlocui. Schema de calcul a mediilor mobile în acest ultim caz (pentru $p = 4$) este prezentată în Tabelul 6.4.

Calculul mediilor mobile dintr-un număr par de termeni

Tabelul 6.4.

Termenii reali ai seriei	Medii mobile parţiale (*MM*)	Medii mobile centrate (Valori ajustate)
y_1		
y_2	$MM_1 = \frac{y_1 + y_2 + y_3 + y_4}{4}$	$\hat{y}_1 = \frac{MM_1 + MM_2}{2} = \frac{\frac{y_1}{2} + y_2 + y_3 + y_4 + \frac{y_5}{2}}{4}$
y_3	$MM_2 = \frac{y_2 + y_3 + y_4 + y_5}{4}$	$\hat{y}_2 = \frac{MM_2 + MM_3}{2} = \frac{\frac{y_2}{2} + y_3 + y_4 + y_5 + \frac{y_6}{2}}{4}$
y_4	$MM_3 = \frac{y_3 + y_4 + y_5 + y_6}{4}$	$\hat{y}_3 = \frac{MM_3 + MM_4}{2} = \frac{\frac{y_3}{2} + y_4 + y_5 + y_6 + \frac{y_7}{2}}{4}$.
y_5	$MM_4 = \frac{y_4 + y_5 + y_6 + y_7}{4}$	
y_6		
y_7		
.....	.....	.....

Se observă că MM_1, MM_2, ... nu apar în dreptul unor termeni reali ai seriei cronologice, deci nu-i pot înlocui, dar mediile aritmetice ale mediilor mobile – da ($\hat{y}_1$ în dreptul lui y_3, $\hat{y}_2$ în dreptul lui y_4 etc.).

Mediile mobile finale, centrate, se pot obţine şi în mod direct, chiar din termenii seriei iniţiale, fără a mai parcurge cele două etape. Astfel, media mobilă finală se determină ca o medie cronologică simplă dintr-un set de $(p + 1)$ termeni (în cazul nostru 5 termeni) iniţiali. Altfel spus, media mobilă centrată se calculează dintr-un număr de $(p + 1)$ termeni, din care primul şi ultimul se însumează cu jumătate din valoarea lor, iar ceilalţi cu întreaga lor valoare. Suma astfel obţinută se împarte la numărul de termeni întregi plus 1.

Şi în acest caz, numărul de medii mobile obţinute este mai mic decât numărul de termeni reali ai seriei. Prima medie mobilă se va plasa în dreptul celui

Scanned with OKEN Scanner

de-al $\frac{p+2}{2}$-lea termen al seriei. Prin această metodă se vor pierde un număr de p termeni reali.

Metoda mediilor mobile se foloseşte în ajustarea îndeosebi a seriilor cronologice afectate de factori sezonieri. Ea prezintă avantajul că este foarte flexibilă, aplicarea ei nu este condiţionată de realizarea unor condiţii prealabile. Are însă şi dezavantajul că prin aplicarea ei se pierde informaţie şi că nu permite prognozarea în viitor a unor noi valori ale fenomenului analizat.

Exemplul 6.6. Vânzările trimestriale realizate de o fabrică de jucării electronice în perioada 2002-2005 se prezintă astfel: (Tabelul 6.5.) (preţuri comparabile – mld. lei):

Vânzările trimestriale în perioada 2002-2005

Tabelul 6.5.

Trimestrul	Anul			
	2002	2003	2004	2005
0	1	2	3	4
I	10	11	14	13
II	14	16	18	16
III	11	10	13	9
IV	21	22	22	25

Datele fiind trimestriale, se vor calcula medii mobile din număr par de termeni ($p = 4$). Mecanismul de calcul şi mediile mobile parţiale şi finale sunt prezentate în Tabelul 6.6.

Tabelul 6.6.

Anul şi trimestrul		Termenii seriei	Medii mobile parţiale	Medii mobile centrate
0		1	2	3
2002	I	10		–
	II	14		–
	III	11	14	14,125
	IV	21	14,25	14,5
2003	I	11	14,75	14,625
	II	16	14,5	14,625
	III	10	14,75	15,125
	IV	22	15,5	15,75
2004	I	14	16	16,375
	II	18	16,75	16,75
	III	13	16,75	16,625
	IV	22	16,5	16,25
2005	I	13	16	15,5
	II	16	15	15,375
	III	9	15,75	–
	IV	25		–

Scanned with OKEN Scanner

c) **Metoda modificării medii absolute**

Foloseşte pentru ajustare, aşa cum îi arată şi numele, un model de calcul ce include modificarea medie absolută:

$$\hat{y}_t = y_1 + (t-1)\overline{\Delta} \quad ,t = \overline{2,n} \tag{6.28.}$$

unde y_1 = primul termen al seriei (termenul de bază);

$\hat{y}_t$ = valoarea ajustată la momentul t.

Metoda poate fi utilizată atunci când modificarea medie absolută determinată are conţinut real şi anume atunci când modificările absolute cu bază în lanţ sunt aproximativ constante, seria formând aproximativ o progresie aritmetică (cu raţia egală cu $\overline{\Delta}$).

Primul şi ultimul termen ajustat, obţinuţi cu această metodă, sunt egali cu termenii reali corespunzători ai seriei cronologice, aşa cum rezultă şi din următoarele relaţii:

$$\hat{y}_1 = y_1 + (1-1)\cdot\overline{\Delta} = y_1, \text{ pentru } t = 1. \tag{6.29.}$$

$$\hat{y}_n = y_1 + (n-1)\overline{\Delta} = y_1 + (n-1)\cdot\frac{y_n - y_1}{n-1} = y_n, \text{ pentru } t = n. \tag{6.30.}$$

Exemplul 6.7. Pentru seria cronologică prezentată în Tabelul 6.1., s-a calculat $\overline{\Delta} = 4{,}3$ pers. Relaţia pe baza căreia se efectuează ajustarea cu metoda modificării medii absolute este:

$$\hat{y}_t = 50 + 4{,}3(t-1), \quad t = \overline{2,n}$$

iar valorile ajustate obţinute sunt prezentate în Tabelul 6.7., coloana 2.

Tabelul 6.7.

Anii	y_t	$\hat{y}_t$ $(\overline{\Delta})$	$\hat{y}_t$ $(\overline{I})$	$(y_t - \hat{y}_t)^2$ $(\overline{\Delta})$	$(y_t - \hat{y}_t)^2$ $(\overline{I})$
0	1	2	3	4	5
1998	50	50	50	0	0
1999	47	54,3	53,5	53,29	42,25
2000	45	58,6	57,24	184,96	149,82
2001	55	62,9	61,25	62,41	39,06
2002	61	67,2	65,54	38,44	20,61
2003	68	71,5	70,13	12,25	4,54
2004	70	75,8	75,04	33,64	25,40
2005	80	80,0	80,0	0	0
TOTAL				384,99	281,68

d) **Metoda indicelui mediu de dinamică**

Este tot o metodă mecanică de determinare a trendului unei serii cronologice, care se aseamănă cu metoda modificării medii absolute, în sensul că

Scanned with OKEN Scanner

modelul de calcul al valorilor trendului include tot un indicator mediu al seriei. Spre deosebire de metoda menționată, acest indicator este indicele mediu de dinamică ($\bar{I}$). Ajustarea se face printr-o relație de forma:

$$\hat{y}_t = y_1 \left(\bar{I}\right)^{t-1}, \quad t = \overline{2,n} \tag{6.31.}$$

Calculul lui $\bar{I}$ are sens (și tot atunci este recomandată a fi folosită și metoda) atunci când toți indicii de dinamică cu bază în lanț sunt aproximativ egali între ei, adică atunci când seria cronologică analizată ia forma unei progresii geometrice cu rația $\bar{I}$.

Și în acest caz, termenii ajustați extremi (primul și ultimul) sunt egali cu termenii reali corespunzători, așa cum rezultă din relațiile (6.32.) și (6.33.).

$$\hat{y}_1 = y_1 \cdot \bar{I}^{(1-1)} = y_1, \text{ pentru } t = 1 \tag{6.32.}$$

$$\hat{y}_n = y_1 \cdot \bar{I}^{n-1} = y_1 \left(\sqrt[n-1]{\frac{y_n}{y_1}} \right)^{n-1} = y_1 \cdot \frac{y_n}{y_1} = y_n, \text{ pentru } t = n \tag{6.33.}$$

Exemplul 6.8. Pentru seria cronologică prezentată în Tabelul 6.1., s-a calculat $\bar{I} = 1{,}07$, iar modelul de ajustare în cazul acestei metode va fi:

$$\hat{y}_t = 50 \cdot 1{,}07^{(t-1)}, \quad t = \overline{2,n}$$

Valorile trendului obținute cu această metodă se regăsesc în coloana 3 a Tabelului 6.7.

Metodele prezentate până acum folosesc pentru calculul valorilor ajustate doar primul și ultimul termen al seriei, neținând cont și de ceilalți termeni intermediari. Dacă sunt satisfăcute condițiile de omogenitate ale modificărilor absolute cu bază în lanț sau ale indicilor de dinamică cu bază în lanț, ele pot oferi informații utile despre evoluția fenomenelor.

6.4.2. Modele analitice de determinare a trendului unei serii cronologice

Metodele analitice oferă o ajustare mai exactă a seriilor cronologice decât cele mecanice, deoarece ele țin cont de toți termenii seriei. Metodele analitice presupun determinarea valorilor de trend cu ajutorul unor funcții analitice (matematice) potrivite, corespunzătoare formei tendinței de evoluție. Problema fundamentală în aplicarea acestor metode este alegerea tipului potrivit de funcție ($f_{(t)}$), numită și **funcție de ajustare a trendului**, care estimează tendința centrală a seriei.

Funcțiile utilizate mai des în practică pentru ajustarea SCR sunt funcțiile matematice uzuale: liniară, polinomială, hiperbolică, parabolică, exponențială, logistică. Alegerea funcției se poate face plecând de la reprezentarea grafică a evoluției fenomenului, de la o analiză complexă a acestui fenomen.

Exemple de funcții de ajustare se prezintă în Figura 6.12.

Scanned with OKEN Scanner

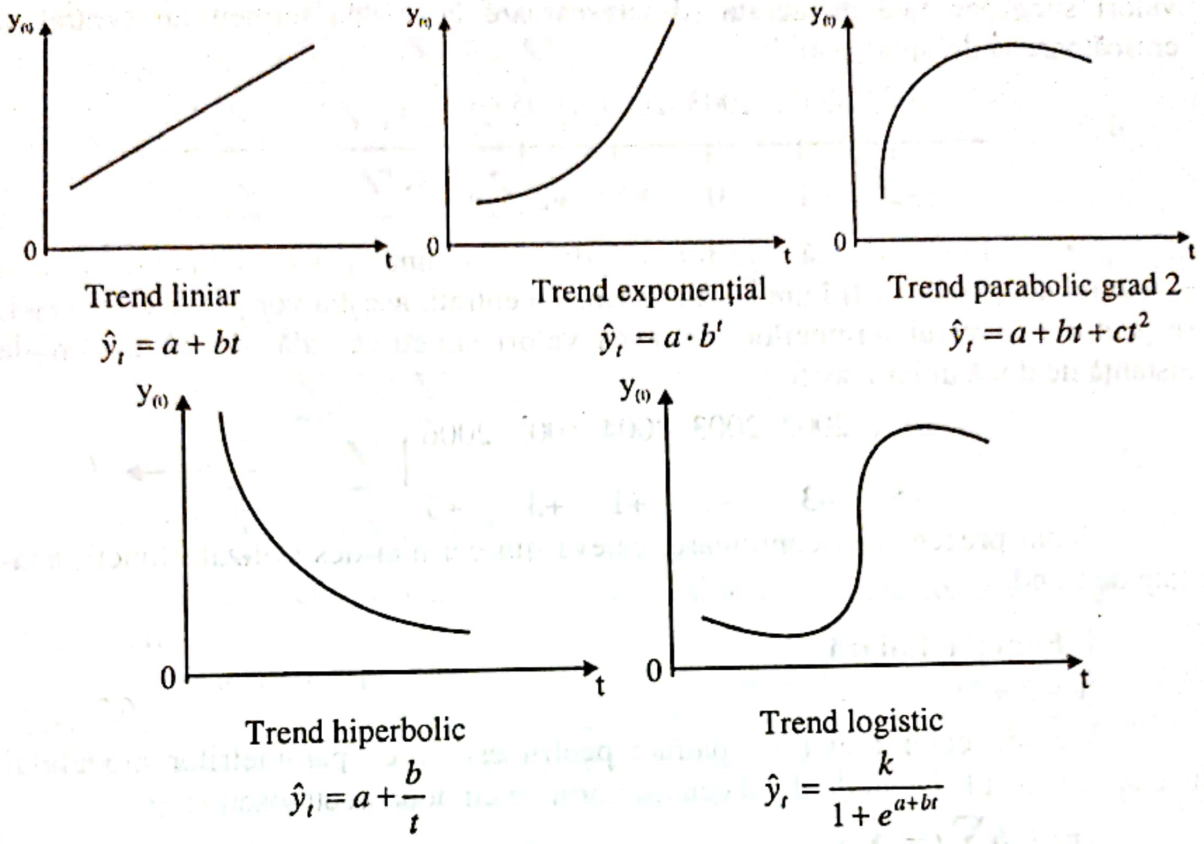


Figura 6.12. Funcțiile de ajustare a seriilor cronologice

Odată aleasă funcția de ajustare, trebuie estimați, în continuare, parametrii. Aceasta se efectuează prin mai multe metode, dintre care cea mai utilizată este metoda celor mai mici pătrate (MCMMP). Metoda presupune ca suma pătratelor abaterilor valorilor ajustate de la cele reale să fie minimă.

$$\sum (y_t - \hat{y}_t)^2 \rightarrow \text{minim} \qquad (6.34.)$$

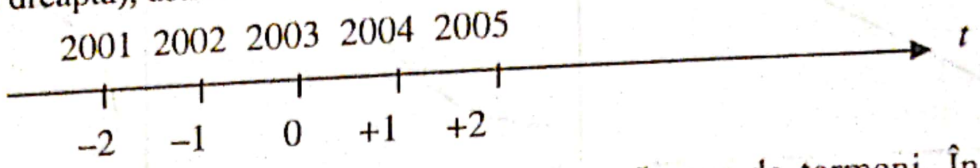
Metoda a fost întâlnită la estimarea parametrilor funcțiilor de regresie (capitolul 5). Deosebirea față de analiza regresiei constă în faptul că locul variabilei independente X este acum luat de variabila timp (t). Nu trebuie interpretat însă că valorile variabilei Y (termenii seriei) depind de succesiunea valorilor lui t, deoarece evoluția fenomenului descris de seria cronologică depinde de o multitudine de factori esențiali și neesențiali, generali și specifici, factori care, și ei, acționează în timp.

În plus, valorile variabilei timp (t) se măsoară cu ajutorul scalei de interval, ce prezintă următoarele proprietăți: originea scalei și unitatea de măsură pot fi alese în mod arbitrar. De aceea, pentru a ușura procesul de calcul, valorile lui t se pot alege convenabil, astfel încât suma lor să fie zero ($\Sigma t = 0$). Se disting două cazuri:

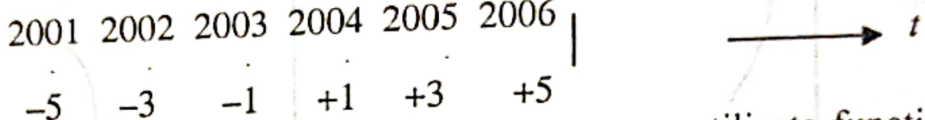
• seria cronologică este formată dintr-un număr impar de termeni. În acest caz, originea ($t = 0$) va fi considerată termenul central, restul termenilor vor avea

Scanned with OKEN Scanner

valori simetrice față de acesta (descrescătoare la stânga termenului central și crescătoare la dreapta), astfel:



• seria cronologică este formată dintr-un număr par de termeni. În acest caz, originea ($t = 0$) va fi între cei doi termeni centrali; aceştia vor primi valorile –1, respectiv +1, restul termenilor vor avea valori simetrice față de cei centrali la distanță de două unități, astfel:



Vom prezenta, în continuare, câteva din cele mai des utilizate funcții analitice de trend:

I. Funcția liniară

$$\hat{y} = a + bt \tag{6.35.}$$

Metoda celor mai mici pătrate pentru estimarea parametrilor modelului duce la următorul sistem de două ecuații normale cu două necunoscute:

$$\begin{cases} na + b\sum t = \sum y_t \\ a\sum t + b\sum t^2 = \sum ty_t \end{cases} \tag{6.36.}$$

Cum $\sum t = 0$, sistemul devine:

$$\begin{cases} na = \sum y_t \\ b\sum t^2 = \sum ty_t \end{cases} \Rightarrow \begin{cases} a = \dfrac{\sum y_t}{n} = \bar{y} \\ b = \dfrac{\sum ty_t}{\sum t^2} \end{cases} \tag{6.37.}$$

Ajustarea după modelul liniar, ca și ajustarea după metoda modificării medii absolute se recomandă dacă evoluția termenilor seriei cronologice urmează aproximativ o progresie aritmetică.

II. Funcția parabolei de gradul doi

$$\hat{y} = a + bt + ct^2 \tag{6.38.}$$

$$\sum (y_t - y_t)^2 \to \min \text{im}$$

$$\sum \left(y_t - a - bt - ct^2\right)^2 \to \min \text{im} \tag{6.39.}$$

Scanned with OKEN Scanner

Rezultă următorul sistem de 3 ecuații cu 3 necunoscute:

$$\begin{cases} na + b\sum t + c\sum t^2 = \sum y_t \\ a\sum t + b\sum t^2 + c\sum t^3 = \sum ty_t \\ a\sum t^2 + b\sum t^3 + c\sum t^4 = \sum t^2 y_t \end{cases} \quad (6.40.)$$

Prin alegerea convenabilă a valorilor de pe axa timpului avem $\sum t = 0$, $\sum t^3 = 0$, iar sistemul devine:

$$\begin{cases} na + c\sum t^2 = \sum y_t \\ b\sum t^2 = \sum ty_t \\ a\sum t^2 + c\sum t^4 = \sum t^2 y_t \end{cases} \quad (6.41.)$$

Prin rezolvarea sistemului se găsesc valorile parametrilor a, b și c și funcția de ajustare.

Ajustarea după modelul parabolic de grad 2 se recomandă atunci când modificările absolute cu bază în lanț alcătuiesc aproximativ o linie dreaptă și diferențele absolute de ordinul 2 (numite și sporuri ale sporurilor) sunt aproximativ constante.

Parabola de gradul 2 are un singur punct de inflexiune (de maxim sau de minim). Dacă linia de evoluție a seriei cronologice are mai multe puncte de inflexiune, se poate folosi, pentru ajustare, o funcție polinomială de grad superior.

III. **Funcția exponențială**

$$\hat{y} = a \cdot b^t \quad (6.42.)$$

Pentru estimarea parametrilor modelului, relația se logaritmează, obținându-se o liniarizare a modelului:

$$\log \hat{y}_t = \log a + t \log b \quad (6.43.)$$

Aplicarea metodei celor mai mici pătrate duce la următorul sistem:

$$\begin{cases} n \log a + \log b \sum t = \sum \log y_t \\ \log a \sum t + \log b \sum t^2 = \sum t \log y_t \end{cases} \quad (6.44.)$$

Cum $\sum t = 0$, sistemul devine:

$$\begin{cases} n \log a = \sum \log y_t \\ \log b \sum t^2 = \sum t \log y_t \end{cases} \Rightarrow \begin{cases} \log a = \dfrac{\sum \log y_t}{n} \\ \log b = \dfrac{\sum t \log y_t}{\sum t^2} \end{cases} \quad (6.45.)$$

Scanned with OKEN Scanner

Prin antilogaritmare se determină a şi b.

Funcţia exponenţială este recomandată a fi utilizată în ajustarea seriei cronologice atunci când termenii acesteia alcătuiesc aproximativ o progresie geometrică (la fel ca în cazul folosirii metodei indicelui mediu de dinamică).

IV. **Funcţia logistică**

Curba logistică a fost introdusă în anul 1845 de matematicianul belgian P.F. Verhulst şi utilizată prima dată în economie în 1930.

Se utilizează în ajustarea unor serii cronologice ai căror termeni înregistrează următoarea evoluţie: în prima perioadă de timp are loc o creştere lentă, care apoi se accelerează exponenţial, până se atinge un punct după care ritmul de creştere încetineşte, tinzând către un nivel limită (nivel de saturaţie), care nu va mai fi depăşit. Relaţia matematică a modelului este:

$$\hat{y}_t = \frac{k}{1+e^{a+bt}} \qquad (6.46.)$$

Exemplul 6.9. Pentru seria cronologică din Tabelul 6.1. se va efectua ajustarea pe baza modelului liniar şi exponenţial. Calculele se prezintă în Tabelul 6.8.

Tabelul 6.8.

Ani	y_t	t	t^2	ty_t	$\hat{y}_t$ (lin.)	$(y_t-\hat{y}_t)^2$	$lg\ y_t$	$t\ lg\ y_t$	$\hat{y}_t$ (exp.)	$(y_t-\hat{y}_t)^2$
0	1	2	3	4	5	6	7	8	9	10
1998	50	–7	49	–350	42,84	51,26	1,69897	–11,89	44,36	31,81
1999	47	–5	25	–235	47,6	0,36	1,672	–8,36	47,98	0,96
2000	45	–3	9	–135	52,36	54,17	1,6532	–4,96	51,90	47,61
2001	55	–1	1	–55	57,12	4,5	1,74036	–1,74	56,13	1,27
2002	61	+1	1	+61	61,88	0,77	1,7853	+ 1,78	60,71	0,08
2003	68	+3	9	+204	66,64	1,85	1,8325	+ 5,50	65,67	5,43
2004	70	+5	25	+350	71,4	1,96	1,8451	+ 9,23	71,02	1,04
2005	80	+7	49	+560	76,16	14,74	1,9031	+13,32	76,82	10,11
Total	476	0	168	400	476	129,61	14,13	2,88		98,31

Pentru modelul liniar:

$$\begin{cases} a = \dfrac{\sum y_t}{n} = \dfrac{476}{8} = 59,5 \approx 60 \text{ pers.} \\ b = \dfrac{\sum ty_t}{\sum t^2} = \dfrac{400}{168} = 2,38 \end{cases}$$

$$\hat{y}_t = 59,5 + 2,38t$$

Valorile ajustate pe baza modelului liniar se regăsesc în coloana 5 a Tabelului 6.8.

Scanned with OKEN Scanner

Pentru modelul exponențial:

$$lg\ a = \frac{\sum lg\ y_t}{n} = \frac{14,13}{8} = 1,76625 \Rightarrow a = 58,378$$

$$lg\ b = \frac{\sum t\ lg\ y_t}{\sum t^2} = \frac{2,88}{168} = 0,017 \Rightarrow b = 1,04$$

$$\hat{y}_t = 58,378 \cdot 1,04^t$$

Valorile ajustate pe baza modelului exponențial sunt calculate în coloana 9 a Tabelului 6.8.

6.4.3. Metode de alegere a celui mai bun model de determinare a tendinței

Am văzut că ajustarea seriilor cronologice se poate face pe baza unui set de metode mecanice și analitice. De aceea este necesară alegerea metodei care asigură o ajustare optimă a seriei.

Alegerea celei mai potrivite metode de ajustare se face pe baza următoarelor criterii:

a) se alege pe reprezentarea grafică a evoluției seriei cronologice acea dreaptă (curbă) trasată pe baza funcției de ajustare care este cea mai apropiată de curba valorilor reale ale seriei;

b) se alege acea funcție de ajustarea a trendului unei serii cronologice pentru care suma valorilor ajustate ($\sum \hat{y}_t$) este cea mai apropiată de suma valorilor reale ($\sum y_t$)

$$\sum \hat{y}_t \approx \sum y_t \qquad (6.47.)$$

Acest criteriu este îndeplinit pentru funcțiile analitice, deoarece se presupune $\sum E_t = 0$.

c) se alege drept optimă acea funcție de ajustare pentru care suma pătratelor diferențelor dintre valorile ajustate și cele reale are valoarea minimă

$$\sum (y_t - \hat{y}_t)^2 \rightarrow \min \text{im} \qquad (6.48.)$$

d) se alege drept optimă pentru ajustarea seriei cronologice acea funcție pentru care abaterea medie pătratică a termenilor ajustați față de cei reali este minimă

$$\sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n}} \rightarrow \min \text{im} \qquad (6.49.)$$

Exemplul 6.10. Aplicând criteriul (c) pentru seria cronologică din Tabelul 6.1. ajustată prin cele patru metode din exemplele anterioare, s-au obținut, pe baza

Scanned with OKEN Scanner

calculelor din Tabelul 6.7., coloanele 4 și 5 și Tabelul 6.8., coloanele 6 și 10, rezultatele:

Metoda de ajustare	$\sum (y_t - \hat{y}_t)^2$
Metoda modificării medii absolute	384,99
Metoda indicelui mediu	281,68
Metoda funcției liniare	129,61
Metoda funcției exponențiale	98,31

Așadar, dintre cele două metode elementare, cea mai potrivită este metoda indicelui mediu, iar dintre cele două metode analitice, mai exactă este metoda trendului exponențial. Dintre toate cele patru metode, metoda trendului exponențial este optimă.

6.5. Modelarea componentei sezoniere a unei serii cronologice

În paragraful 6.3. s-a arătat că una din componentele seriilor cronologice este componenta sezonieră. Atunci când valorile reale ale unui fenomen se abat periodic de la tendința generală (trend), cauza poate fi acțiunea unor factori sezonieri. Scopul unei analize statistice este de a determina momentele sau perioadele în care se înregistrează această abatere, direcția și amplitudinea variației.

De cele mai multe ori, oscilațiile sezoniere apar ca urmare a influenței (succesiunii) anotimpurilor asupra fenomenelor. Factorul sezonier afectează într-o mai mare măsură nivelul fenomenelor din sfera turismului, industriei alimentare (de conserve), de energie electrică, producției de jucării, de rechizite școlare etc.

Determinarea oscilațiilor sezoniere depinde de modelul ales ce leagă toate componentele seriei cronologice. În vederea obținerii oscilațiilor sezoniere, trebuie eliminate din nivelul real al termenilor seriei celelalte componente (trendul, componenta aleatorie și cea ciclică). Dacă seria cronologică este staționară, deci nu prezintă trend, atunci trebuie eliminată doar componenta aleatorie și cea ciclică. De cele mai multe ori, însă, nu se dispune de date suficiente pentru identificarea componentei ciclice, astfel că cel mai adesea ea este necunoscută.

Determinarea componentei sezoniere ar putea facilita desfășurarea activității de producție și economisirea resurselor financiare în cadrul agenților economici și ar ține sub control efectele factorilor sezonieri.

a) Dacă **influența factorului sezonier se manifestă aditiv** (se folosește modelul aditiv de compunere a elementelor seriei cronologice), atunci componenta sezonieră se determină sub forma devierilor sezoniere (S_j).

Pentru a evalua abaterile sezoniere, este necesară determinarea prealabilă a trendului cu una din metodele prezentate anterior (o metodă analitică sau metoda

Scanned with OKEN Scanner

mediilor mobile – ultima fiind mai indicată atunci când seria cronologică este afectată de factori sezonieri).

Algoritmul pentru determinarea devierilor sezoniere cuprinde următorii paşi:

- se elimină din termenii reali ai seriei trendul (prin scădere):

$$y_{ij} - \hat{y}_{ij} = \left(\hat{y}_{ij} + S_j + \varepsilon_{ij}\right) - \hat{y}_{ij} = S_j + \varepsilon_{ij} \qquad (6.50.)$$

unde $i = \overline{1,n}$ (nr. curent al perioadei)

$j = \overline{1,m}$ (nr. curent al subperioadei – sezonului)

- se elimină şi influenţa factorului aleatoriu, prin calculul unor medii ale diferenţelor calculate la pasul anterior, pentru fiecare sezon (j):

$$S'_j = \frac{\sum_{i=1}^{n}\left(y_{ij} - \hat{y}_{ij}\right)}{n} = \frac{\sum_{i=1}^{n}\left(S_j + \varepsilon_{ij}\right)}{n} = S_j + \frac{\sum_{i=1}^{n}\varepsilon_{ij}}{n} \approx S_j \qquad (6.51.)$$

Valorile S'_j determinate reprezintă estimatori bruţi ai componentei sezoniere. Dacă trendul a fost determinat cu MCMMP, suma abaterilor sezoniere este nulă $\left(\sum_{j=1}^{m} S'_j = 0\right)$ şi deci media lor este nulă, întrucât

$$\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} \hat{y}_{ij} \qquad (6.52.)$$

Dacă însă trendul a fost determinat cu metoda mediilor mobile, compensarea abaterilor sezoniere nu are loc în mod obligatoriu şi se trece la pasul următor.

- din estimatorii bruţi ai devierilor sezoniere se calculează medii ale acestora

$$\frac{\sum_{j=1}^{m} s'_j}{m} \qquad (6.53.)$$

Se observă că la pasul anterior am obţinut m estimatori bruţi ai devierilor sezoniere (de exemplu, dacă am înregistrat vânzările trimestriale de bilete ale unei agenţii de turism, pentru trei ani consecutivi, obţinem, pentru fiecare trimestru, 3 valori diferite ale abaterilor sezoniere, corespunzătoare celor 3 ani în care s-au înregistrat datele). De aceea se caracterizează fiecare trimestru printr-o singură valoare, medie, a abaterilor sezoniere brute.

- mediile obţinute la pasul anterior se scad din devierile sezoniere brute, obţinându-se devierile sezoniere corectate (S_j):

$$S_j = S'_j - \frac{\sum_{j=1}^{m} S'_j}{m} = \frac{\sum_{j=1}^{m}\left(y_{ij} - \hat{y}_{ij}\right)}{m} = \frac{\sum_{j=1}^{m} S'_j}{m} \qquad (6.54.)$$

Scanned with OKEN Scanner

Astfel, se obţin m devieri sezoniere. Valoarea devierii sezoniere corectate S_j pentru sezonul j se interpretează astfel: în fiecare sezon j, componenta sezonieră deviază termenii seriei cronologice cu $|S_j|$ de la valoarea de trend. Semnul abaterii sezoniere indică direcţia în care s-a produs abaterea sezonieră (plus – înseamnă că factorul sezonier a determinat o creştere, minus – înseamnă că factorul sezonier a determinat o scădere a termenilor reali de la trend)

• se determină termenii seriei cronologice corectate (adică termenii din care s-a eliminat influenţa factorului sezonier):

$$y_{ij} - S_j, \qquad j = \overline{1,m} \tag{6.55.}$$

Seria corectată de sezonalitate va include doar trendul şi abaterile aleatorii

$$y_{ij} - S_j = \hat{y}_{ij} + \varepsilon_{ij}, \qquad \begin{matrix} i = \overline{1,n} \\ j = \overline{1,m} \end{matrix} \tag{6.56.}$$

Dacă se cunoaşte trendul, se poate calcula pentru fiecare termen, valoarea componentei aleatorii:

$$y_{ij} - \hat{y}_{ij} - S_j = \varepsilon_{ij} \tag{6.57.}$$

b) Dacă **influenţa factorului sezonier se manifestă multiplicativ** (se foloseşte modelul multiplicativ de compunere a elementelor seriei cronologice), atunci componenta sezonieră ce se va determina îmbracă forma indicilor de sezonalitate (S_j^*).

Algoritmul pentru calculul indicilor de sezonalitate cuprinde următorii paşi:

• se elimină (prin împărţire) valoarea trendului, din termenii reali:

$$\frac{y_{ij}}{\hat{y}_{ij}} = S_j^* \cdot \varepsilon_{ij}^* \tag{6.58.}$$

Ceea ce rămâne în urma împărţirii este componenta sezonieră şi cea aleatorie, legate prin relaţia de produs.

• se calculează medii pe subperioade (sezoane) ale acestor rapoarte, eliminându-se astfel influenţa factorului aleatoriu, mărimilor obţinute numindu-se indici de sezonalitate „bruţi".

$$S_j^{*'} = \frac{\sum_{i=1}^{n} \frac{y_{ij}}{\hat{y}_{ij}}}{n} = \frac{\sum_{i=1}^{n} S_j^* \cdot \varepsilon_{ij}^*}{n} = S_j^* \cdot \frac{\sum_{i=1}^{n} \varepsilon_{ij}^*}{n} \approx S_j^* \tag{6.59.}$$

Este adevărat că, teoretic, pentru calculul indicilor bruţi de sezonalitate ar trebui aplicată media geometrică, însă, în practică, se recurge, de obicei, la media aritmetică:

Scanned with OKEN Scanner

• dacă produsul indicilor bruți de sezonalitate diferă de 1 (sau 100%), atunci se determină indici de sezonalitate corectați, prin împărțirea indicilor bruți la media lor, calculată pentru un anumit sezon.

$$S_j^* = \frac{S_j^{*'}}{\frac{\sum_{j=1}^{m} S_j^{*'}}{m}} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{y_{ij}}{\hat{y}_{ij}}}{\frac{1}{m}\sum_{j=1}^{m} S_j^{*'}} \qquad (6.60.)$$

Se obțin, astfel, m indici de sezonalitate. Semnificația lui $|S_j^*|$ este aceea că, în fiecare sezon, factorul sezonier deviază valorile reale ale termenilor seriei de S_j^* ori de la linia de trend, sau cu $(S_j^* - 1)100$ procente. Atunci se poate determina componenta aleatorie:

$$\frac{y_{ij}}{\hat{y}_{ij} \cdot S_j^*} = \varepsilon_{ij}^* \qquad (6.61.)$$

Seria cronologică corectată de sezonalitate (desezonalizată) se obține împărțind termenii reali ai seriei la indicii de sezonalitate:

$$\frac{y_{ij}}{S_j^*} = \hat{y}_{ij} \cdot \varepsilon_{ij}^* \qquad (6.62.)$$

Metodele descrise se aplică după ce a fost verificată, în prealabil, omogenitatea datelor.

Seriile de timp corectate (desezonalizate) au rolul de a arăta cum ar evolua valorile de trend, dacă nu ar fi influențate de factorii sezonieri. Ele păstrează, totuşi, influența factorilor ciclici şi aleatorii.

Constituirea unor astfel de serii corectate se bazează pe identificarea corectă a factorilor sezonieri, pornind de la datele inițiale, care se referă, desigur, la o perioadă trecută. Pe baza modelelor adecvate, se pot efectua şi previzionări pentru o perioadă viitoare.

Totuşi, este necesară o anumită precauție în utilizarea acestor modele, întrucât factorii sezonieri, identificați pe baza datelor despre o perioadă trecută ar putea să-şi modifice în viitor acțiunea asupra fenomenului. De aceea, aceste modele pot forma o soluție acceptabilă pe termen scurt, iar influența factorilor sezonieri ar trebui recalculată pe măsură ce apar noi date disponibile.

Exemplul 6.11. Se vor folosi datele din Tabelul 6.5., pentru care s-au calculat mediile mobile din Tabelul 6.6. Pentru analiza sezonalității vom aplica, pe rând, modelul aditiv şi cel multiplicativ.

Scanned with OKEN Scanner

a) **Modelul aditiv**

• se elimină din termenii reali y_t valorile de trend $(\hat{y}_t)$, determinate prin metoda mediilor mobile. Rezultatele se găsesc în coloana 3 a Tabelului 6.9., începând cu trimestrul III/2002:

$(y_t - \hat{y}_t)_{III/02} = 11 - 14{,}125 = -3{,}125$ pers.

$(y_t - \hat{y}_t)_{IV/02} = 21 - 14{,}5 = 6{,}5$ pers.

etc.

• aceste diferențe se trec în Tabelul 6.10., în care se calculează devierile sezoniere brute (S'_j) – coloana 5 și corectate (s_j) – coloana 6.

Devierile sezoniere brute:

$$S_I' = \frac{-3{,}125 - 2{,}375 - 2{,}5}{3} = -2{,}667 \text{ pers.}$$

$$S_{II}' = \frac{1{,}375 + 1{,}25 + 0{,}625}{3} = 1{,}083 \text{ pers.}$$

$$S_{III}' = \frac{-3{,}125 - 5{,}125 - 3{,}625}{3} = -3{,}958 \text{ pers.}$$

$$S_{IV}' = \frac{6{,}5 + 6{,}25 + 5{,}75}{3} = 6{,}167 \text{ pers.}$$

Media devierilor sezoniere brute:

$$\overline{S'} = \frac{S_I' + S_{II}' + S_{III}' + S_{IV}'}{4} = \frac{-2{,}667 + 1{,}083 - 3{,}958 + 6{,}167}{4} = 0{,}15625$$

Devierile sezoniere corectate:

$$S_I = S_I' - \overline{S'} = -2{,}667 - 0{,}15625 = -2{,}823 \cong -3 \text{ pers.}$$

$$S_{II} = S_{II}' - \overline{S'} = 1{,}083 - 0{,}15625 = 0{,}927 \cong 1 \text{ pers.}$$

$$S_{III} = S_{III}' - \overline{S'} = -3{,}958 - 0{,}15625 = -4{,}114 \cong -4 \text{ pers.}$$

$$S_{IV} = S_{IV}' - \overline{S'} = 6{,}167 - 0{,}15625 = 6{,}011 \cong 6 \text{ pers.}$$

• se calculează termenii seriei cronologice corectate, eliminându-se din valorile termenilor reali devierile sezoniere corectate (coloana 4, Tabelul 6.9):

$(y_t - S_I)_{II/02} = 10 - (-3) = 13$ pers.

$(y_t - S_{II})_{III/02} = 14 - 1 = 13$ pers.

etc.

• se determină componenta aleatoare, scăzând din valorile termenilor reali trendul și sezonabilitatea (coloana 5, Tabelul 6.9).

Scanned with OKEN Scanner

$(\varepsilon_t)_{III'02} = 11 - 14{,}125 + 4 = 0{,}989$

$(\varepsilon_t)_{IV'02} = 21 - 14{,}5 - 6 = 0{,}489$

etc.

Tabelul 6.9.

Anul şi trimestrul		y_t	Trendul (medii mobile) $\hat{y}_t$	$y_t - \hat{y}_t$	SCR corectată $y_t - s_j$	Componenta aleatorie ε_t
0		1	2	3	4	5
2002	I	10	–	–	13	–
	II	14	–	–	13	–
	III	11	14,125	–3,125	15	0,989
	IV	21	14,5	6,5	15	0,489
2003	I	11	14,625	–3,125	14	–0,802
	II	16	14,625	1,375	15	0,448
	III	10	15,125	–5,125	14	–1,011
	IV	22	15,75	6,25	16	0,239
2004	I	14	16,375	–2,375	17	0,448
	II	18	16,75	1,25	17	0,323
	III	13	16,625	–3,625	17	0,489
	IV	22	16,25	5,75	16	–0,261
2005	I	13	15,5	–2,5	16	0,323
	II	16	15,375	0,625	15	–0,302
	III	9	–	–	13	–
	IV	25	–	–	19	–

Tabelul 6.10.

Trim.	Diferenţe sezoniere				Devieri sezoniere brute (S_j')	Devieri sezoniere corectate (S_j)
	2002	2003	2004	2005		
0	1	2	3	4	5	6
I	–	–3,125	–2,375	–2,5	–2,667	–2,823 ≃ –3
II	–	1,375	1,25	0,625	1,083	0,927 ≃ 1
III	–3,125	–5,125	–3,625	–	–3,958	–4,114 ≃ –4
IV	6,5	6,25	5,75	–	6,167	6,011 ≃ 6
	Total				$\overline{S'} = 0{,}15625$	

Scanned with OKEN Scanner

b) **Modelul multiplicativ**

• se împart termenii reali (y_t) la trend ($\hat{y}_t$) (coloana 3, Tabelul 6.11.)

$(y_t - \hat{y}_t)_{III/02} = 11 : 14,125 = 0,779$

$(y_t - \hat{y}_t)_{IV/02} = 21 : 14,5 \quad = 1,448$

etc.

• rapoartele rezultate la pasul anterior se trec în Tabelul 6.12., în care se calculează indicii de sezonalitate bruți (coloana 5) și cei corectați (coloana 6).

Indicii de sezonalitate bruți:

$$S_I^{*'} = \frac{0,779 + 0,855 + 0,839}{3} = 0,824$$

$$S_{II}^{*'} = \frac{1,094 + 1,075 + 1,041}{3} = 1,07$$

$$S_{III}^{*'} = \frac{0,779 + 0,661 + 0,792}{3} = 0,741$$

$$S_{IV}^{*'} = \frac{1,448 + 1,397 + 1,354}{3} = 1,400$$

Media indicilor de sezonalitate bruți:

$$\overline{S}^{*'} = \frac{S_I^{*'} + S_{II}^{*'} + S_{III}^{*'} + S_{IV}^{*'}}{4} = \frac{0,824 + 1,07 + 0,741 + 1,400}{3} = 1,00875$$

Indicii de sezonalitate corectați:

$$S_I^* = \frac{S_I^{*'}}{\overline{S}^{*'}} = \frac{0,824}{1,00875} = 0,817$$

$$S_{II}^* = \frac{S_{II}^{*'}}{\overline{S}^{*'}} = \frac{1,07}{1,00875} = 1,061$$

$$S_{III}^* = \frac{S_{III}^{*'}}{S^{*'}} = \frac{0,741}{1,00875} = 0,734$$

$$S_{IV}^* = \frac{S_{IV}^{*'}}{\overline{S}^{*'}} = \frac{1,4}{1,00875} = 1,388$$

• se calculează termenii seriei cronologice corectate, împărțindu-se termenii reali la indicii de sezonalitate bruți (coloana 4, Tabelul 6.11.);

• se determină componenta aleatorie, împărțindu-se termenii reali la produsul dintre trend și indicii de sezonalitate (coloana 5, Tabelul 6.11.).

Scanned with OKEN Scanner

Tabelul 6.11.

Anul şi trim.		y_t	Trendul (medii mobile) $\hat{y}_t$	$y_t / \hat{y}_t$	SCR corectată y_t / s_j^*	Componenta aleatorie ε_t^*
0	1	2	3	4	5	6
2002	I	10	–	–	12,240 ≃ 12	–
	II	14	–	–	13,195 ≃ 13	–
	III	11	14,125	0,779	14,986 ≃ 15	1,061
	IV	21	14,5	1,448	15,130 ≃ 15	1,043
2003	I	11	14,625	0,779	13,464 ≃ 13	0,921
	II	16	14,625	1,094	15,080 ≃ 15	1,031
	III	10	15,125	0,661	13,624 ≃ 14	0,901
	IV	22	15,75	1,397	15,850 ≃ 16	1,006
2004	I	14	16,375	0,855	17,136 ≃ 17	1,046
	II	18	16,75	1,075	16,965 ≃ 17	1,013
	III	13	16,625	0,782	17,711 ≃ 18	1,065
	IV	22	16,25	1,354	15,850 ≃ 16	0,975
2005	I	13	15,5	0,839	15,912 ≃ 16	1,026
	II	16	15,375	1,041	15,080 ≃ 15	0,981
	III	9	–	–	12,261 ≃ 12	–
	IV	25	–	–	18,011 ≃ 18	–

Tabelul 6.12.

Trim.	Indici de sezonalitate				Indici de sezonalitate bruți $S_j^{*'}$	Indici de sezonalitate corectați (S_j^*)
	2002	2003	2004	2005		
0	1	2	3	4	5	6
I	–	0,779	0,855	0,839	0,824	0,817
II	–	1,094	1,075	1,041	1,07	1,061
III	0,779	0,661	0,782	–	0,741	0,734
IV	1,448	1,397	1,354	–	1,400	1,388
	T o t a l				1,00875	

6.6. Analiza componentei ciclice a unei serii cronologice

Am văzut că una din componentele termenilor reali ai unei serii cronologice este componenta ciclică. Deşi se poate vorbi de existenţa unor factori şi se poate încerca identificarea lor prin utilizarea unor date anuale disponibile, identificarea şi măsurarea acestor variaţii ciclice este dificil de realizat, deoarece aceste

Scanned with OKEN Scanner

cicluri au rareori o lungime predefinită. O altă problemă ar fi aceea că am avea nevoie de 6 sau 7 cicluri complete de date anuale, pentru a identifica în mod cert existența acestor factori ciclici și dimensiunea efectelor lor. De cele mai multe ori nici nu se dispune de serii cronologice atât de lungi.

De-a lungul timpului, au fost identificate câteva cicluri în economie, dintre care amintim:

- *marele val Kondratieff* – în anii 1920, cu o lungime de 40-60 de ani, însă nu există suficiente date pentru a releva cu certitudine existența acestui ciclu;
- *marele val Kuznets* – un ciclu cu o lungime de aproximativ 20 de ani, pentru evidențierea sa s-au utilizat unele studii demografice (migrație);
- *building cycle* – s-a remarcat în mai multe țări, având o lungime de 15-20 de ani;
- *cicluri de afaceri* – nu sunt periodice și au o lungime variabilă (între 1 și 12 ani);
- *cicluri Hansen* – pot fi minime (2-4 ani lungime) sau maxime (6-11 ani lungime).

O modalitate de identificare a influenței factorilor ciclici asupra unei serii cronologice este aceea de a construi graficul de evoluție a datelor anuale și de a căuta modul în care s-ar potrivi cu unul din ciclurile descrise mai sus.

6.7. Analiza componentei aleatorie a unei serii cronologice

În paragraful 6.3. s-au făcut o serie de presupuneri asupra oscilațiilor determinate de factorii aleatorii, în calitatea lor de componente ale seriilor cronologice: media componentei sezoniere este zero în cazul modelului aditiv și unu în cazul modelului multiplicativ.

În exemplul 8 s-au calculat valorile componentei aleatorii pentru modelul aditiv (coloana 5, Tabelul 6.9.) și pentru cel multiplicativ (coloana 5, Tabelul 6.11.). Pe aceste rezultate se pot verifica presupunerile făcute în cazul componentei aleatorii.

Astfel, dacă se utilizează modelul aditiv, media oscilațiilor sezoniere pentru datele din exemplul (6.11.) este de 0,114, care nu diferă semnificativ de zero, iar dacă se utilizează modelul multiplicativ se obține o medie a oscilațiilor sezoniere de 1,00575, care este apropiată de 1. Așadar, presupunerile se adeveresc.

Pe termen lung, ar fi necesară și observarea valorilor pozitive și negative ale componentei sezoniere (în cazul modelului aditiv), respectiv observarea valorilor sub și supraunitare (în cazul modelului multiplicativ), întrucât acestea pot indica existența influenței unor factori ciclici în cadrul modelului.

Scanned with OKEN Scanner

6.8. Covariația. Covariație cu întârziere (decalaj)

Dintre evoluțiile numeroaselor fenomene economico-sociale, unele prezintă un interes suplimentar atunci când sunt analizate comparativ – două sau mai multe serii cronologice. Putem exemplifica, astfel: evoluția prețului unui produs – privită comparativ cu evoluția cantității vândute; evoluția importului în paralel cu cea a exportului etc. Altfel spus, este necesară analiza dependenței, a legăturii dintre evoluțiile în timp a două fenomene. În acest caz, variabilele luate în studiu sunt:

– variabila cauzală, exogenă, factorială (X_t);

– variabila rezultativă, endogenă, dependentă (Y_t);

– variabila timp (t).

Spre deosebire de cele studiate în capitolul 5, aici, între variabila „cauză" și variabila „efect" nu există, de fapt, o legătură reală, ci una care ar putea fi considerată, mai degrabă, „artificială". Măsurarea intensității unei astfel de legături se realizează prin indicatorul numit **covariație**.

O modalitate de exprimare a covariației este coeficientul de covariație liniară, ce se calculează asemănător cu coeficientul de corelație, însă cu conținut diferit. Relația de calcul este:

$$C = \frac{\sum \frac{X_t - \overline{X}}{\sigma_{\overline{x}}} \cdot \frac{Y_t - \overline{Y}}{\sigma_{\overline{y}}}}{n} = \frac{\sum (X_t - \overline{X})(Y_t - \overline{Y})}{\sqrt{\sum (X_t - \overline{X})^2 (Y_t - \overline{Y})^2}} \qquad (6.63.)$$

unde $\frac{X_t - \overline{X}}{\sigma_{\overline{x}}}$ și $\frac{Y_t - \overline{Y}}{\sigma_{\overline{y}}}$ sunt variabilele centrate și reduse.

Coeficientul de covariație liniară măsoară intensitatea legăturii variațiilor în timp dintre fenomene, luând valori între –1 și +1. Dacă valorile sale sunt apropiate de ± 1, atunci legătura liniară dintre variațiile temporale ale celor două fenomene este puternică, iar dacă valorile coeficientului tind spre (0), legătura este slabă.

În anumite cazuri, poate exista legătură între evoluțiile în timp a două variabile (cauză și efect), dar între aceste evoluții să existe din decalaj, un interval. Altfel spus, cele două evoluții „se corelează", dar după un interval de timp (Figura 6.13.).

Scanned with OKEN Scanner

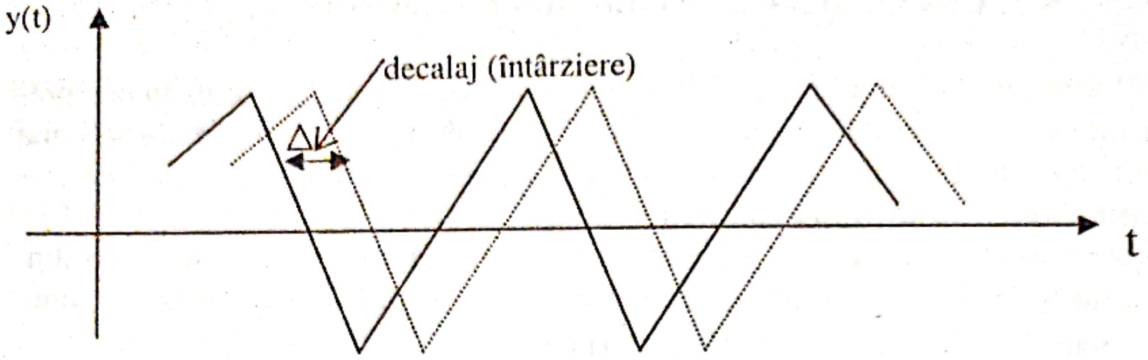


Figura 6.13. Decalajul dintre evoluţiile a două variabile în timp

6.9. Previzionarea pe baza seriilor cronologice

Pentru efectuarea previziunilor fenomenului reflectat de o serie cronologică, se impune recombinarea componentelor seriei, după ce acestea au fost în prealabil identificate şi măsurate, aşa cum s-a arătat în paragrafele anterioare.

Determinarea valorilor prognozate ale fenomenului analizat în viitor presupune prelungirea (extinderea) trendului în perioada următoare perioadei de analiză. Pentru aceasta se alege un model optim de ajustare ($Y_t = f(t)$), şi apoi se introduc în acest model valorile „prelungite" ale variabilei timp (t'), corespunzătoare momentului pentru care se face previziunea. Mai dificilă este „prelungirea" trendului, dacă el s-a determinat cu metoda mediilor mobile. Dacă totuşi trendul s-a determinat prin această metodă, este indicată folosirea unor modele autoregresive (metoda Box-Jenkins).

Aplicarea procedeului menţionat mai sus pentru efectuarea prognozei presupune că factorii, condiţiile care au acţionat asupra fenomenului în orizontul analizat se menţin şi în orizontul de prognoză. Totodată, pentru sesizarea tuturor regularităţilor manifestate în evoluţia fenomenului este necesară efectuarea analizei seriei cronologice pe un interval, pe o perioadă destul de mare de timp, şi seria să aibă un număr suficient de mare de termeni.

De asemenea, orizontul de prognoză nu ar trebui să fie prea mare, recomandându-se în practică ca el să aibă o lungime maximă egală cu o treime din lungimea intervalului pentru care s-a determinat tendinţa generală.

În funcţie de modelele de ajustare, extrapolarea se poate face pe seama următoarelor relaţii:

– *în cazul ajustării cu metode mecanice:*

$$\hat{Y}_t = y_1 + (t'-1)\overline{\Delta} \quad \text{(metoda modificării medii absolute)} \qquad (6.64.)$$

Scanned with OKEN Scanner

$t' = \overline{n+1, n+k}$ = (orizont de prognoză)

$\hat{Y}_{t'} = y(I)^{t'-1}$ (metoda indicelui mediu), (6.65.)

$t' = \overline{n+1, n+k}$

– *în cazul ajustării cu metode analitice:*

$\hat{Y}_{t'} = a + bt'$ pentru trend liniar (6.66.)

$\hat{Y}_{t'} = a + bt' + ct'^2$ pentru trend parabolic de gradul 2 (6.67.)

În cazul seriilor cronologice afectate de oscilaţii sezoniere (care cuprind date trimestriale, lunare etc.), valoarea extrapolată pentru anul k şi sezonul j se determină:

– *pentru modelul aditiv:*

$$Y'_{kj} = \hat{Y}_{kj} \cdot S^*_j \quad (6.68.)$$

unde $\hat{Y}_{kj}$ = valoarea de trend, determinată prin altă metodă decât metoda mediilor mobile, pentru anul k şi sezonul J.

S_j = devierea sezonieră pentru sezonul j

– *pentru modelul multiplicativ:*

$$\hat{Y}'_{kj} = \hat{Y}_{kj} + S^*_j \quad (6.69.)$$

unde S^*_j = indicele de sezonalitate al sezonului j.

Referitor la prognoză, se mai poate răspunde la următoarele probleme:

a) *Care va fi nivelul fenomenului studiat peste K ani?*

Vom răspunde la această problemă efectuând întâi, prognoza, pe baza sporului mediu absolut, apoi, într-o a doua variantă, pe baza indicelui mediu de dinamică.

- **În primul caz** – se presupune că fenomenul evoluează aproximativ într-o progresie aritmetică (deci modificările absolute anuale cu bază în lanţ sunt aproximativ egale, şi ele evoluează liniar).

Fie un fenomen y care, într-o perioadă anterioară, a crescut cu un spor mediu absolut anual $\overline{\Delta}$. Se cere să se determine nivelul pe care-l va atinge fenomenul peste k ani. Dacă presupunem că şi peste k ani fenomenul se va dezvolta în progresie aritmetică şi cu raţia $\overline{\Delta}$, atunci peste k ani fenomenul va avea nivelul dat de relaţia:

$$y_k = y_1 + (k-1)\,\overline{\Delta} \quad (6.70.)$$

Scanned with OKEN Scanner

• **În al doilea caz** – se presupune că fenomenul evoluează aproximativ într-o progresie geometrică (deci indicii de dinamică cu bază în lanț sunt aproximativ egali).

Fenomenul y a evoluat în perioada anterioară într-o progresie geometrică cu rația $\bar{I}$ și se presupune că își va menține această evoluție și în următorii ani, atunci nivelul fenomenului peste k ani va fi:

$$y_k = y_1 \cdot \bar{I}^{k-1} \qquad (6.71.)$$

b) *După câți ani, fenomenul studiat va atinge un anumit nivel, prestabilit?*

• **În cazul folosirii modelului modificării medii absolute**

Să presupunem că, potrivit programului de dezvoltare a unei regiuni, din România, este necesar ca volumul activității să se dubleze. Întrebarea cere să aflăm după câți ani (t) se va întâmpla acest lucru, presupunând că evoluția fenomenului în viitor se păstrează în aceleași condiții ca și în trecut (când a evoluat cu o modificare medie absolută de $\bar{\Delta}$).

$$y_n = 2y_k, \quad n-k=t \qquad (6.72.)$$

$$y_n = y_k + (n-k)\cdot\bar{\Delta} \qquad (6.73.)$$

Substituim $y_n = 2y_k$ și $n-k=t$ în relația (6.73.), rezultând:

$$2y_k = y_k + t\bar{\Delta}$$

$$y_k = t\bar{\Delta} \Rightarrow \quad t = \frac{y_k}{\bar{\Delta}} \qquad (6.74.)$$

• **În cazul folosirii modelului bazat pe indicele mediu de dinamică**

Presupunem că fenomenul a evoluat în trecut în progresie geometrică de rație $\bar{I}$ și că în viitor el va evolua în aceleași condiții. Vom avea:

$$y_t = y_0 \cdot \bar{I}^t \qquad (6.75.)$$

unde y_0 = nivelul fenomenului la data efectuării prognozei.

Dar $y_t = 2y_0$,

iar $\bar{I}^t = 2$

$$lg\,2 = lg\,\bar{I} \Rightarrow t = \frac{lg\,2}{lg\,\bar{I}} \qquad (6.76.)$$

Scanned with OKEN Scanner

c) *Luându-se în considerare două zone, A şi B, să se determine după cât timp se vor egaliza nivelurile de dezvoltare ale celor două zone?*

• În cazul utilizării metodei modificării medii absolute

Presupunem că o zonă *A* îşi propune să ajungă din urmă o altă zonă, *B*, mai dezvoltată din punctul de vedere al unui indicator *y*. Se cunoaşte că fenomenul *y* a crescut cu un spor mediu anual absolut $\bar{\Delta}_A$, respectiv $\bar{\Delta}_B$. Se cere să aflăm după câţi ani *t*, zona *A* va ajunge din urmă zona *B*?

La data când se efectuează prognoza, fenomenul avea nivelul y_o^A, respectiv y_o^B.

În momentul egalizării nivelului indicatorului, vom avea relaţia:

$$y_t^A = y_o^A + t\bar{\Delta}^A \qquad (6.77.)$$

$$y_t^B = y_o^B + t\bar{\Delta}^B \qquad (6.78.)$$

Cum $y_t^A = y_t^B \Rightarrow y_o^A + t\bar{\Delta}^A = y_o^B + t\bar{\Delta}^B \Rightarrow t\left(\bar{\Delta}^A - \bar{\Delta}^B\right) = y_o^B - y_o^A$

$$t = \frac{y_o^B - y_o^A}{\bar{\Delta}_A - \bar{\Delta}_B} \qquad (6.79.)$$

• În cazul utilizării metodei indicelui mediu

Aceeaşi problemă, cu deosebirea că fenomenul *y* s-a dezvoltat în trecut cu indicii medii I_A, I_B şi avea nivelul y_o^A, y_o^B în momentul efectuării prognozei.

Se presupune că $\bar{I}_A > \bar{I}_B$

În momentul egalizării nivelului fenomenului, vom avea relaţia:

$$y_o^A \cdot \bar{I}_A^t = y_o^B \cdot \bar{I}_B^t \qquad (6.80.)$$

Pentru determinarea lui *t* se logaritmează dubla egalitate:

$$lgy_o^A + tlg\bar{I}_A = lgy_o^B + tlg\bar{I}_B$$

$$t\left(lg\bar{I}_A - lg\bar{I}_B\right) = lgy_o^B - lgy_o^A \qquad (6.81.)$$

$$\Rightarrow \quad t = \frac{lgy_o^B - lgy_o^A}{lg\bar{I}_A - lg\bar{I}_B}$$

d) *Determinarea gradului în care o regiune devansează altă regiune*

Se calculează coeficienţii (indicii) de devansare, în două ipostaze:

– pentru două unităţi teritoriale diferite;

– pentru două variabile ce evoluează în timp, dar în interiorul aceleiaşi colectivităţi.

Scanned with OKEN Scanner

Coeficientul de devansare rezultă din raportul a doi indici de dinamică calculați pentru două unități teritoriale diferite, şi arată de câte ori creşte mai repede nivelul fenomenului în unitatea raportată, față de evoluția nivelului fenomenului în unitatea teritorială bază de raportare.

$$I^{y}_{A/B} = \frac{I_A}{I_B} \qquad (6.82.)$$

unde A, B – cele două unități teritoriale

$$I_A = \frac{y^A_n}{y^A_1} \qquad \text{este indicele raportat} \qquad (6.83.)$$

$$I_B = \frac{y^B_n}{y^B_1} \qquad \text{este indicele bază de raportare} \qquad (6.84.)$$

Scanned with OKEN Scanner

CAPITOLUL 7
INDICII STATISTICI

În practică se întâlnesc adesea forme de analiză comparativă a unor fenomene, înregistrate în unităţi de timp sau spaţiu diferite. Astfel, spunem că în anul 2006 judeţul Constanţa a realizat o producţie agricolă de 1,5 ori mai mare decât judeţul Giurgiu; sau că în anul 2005 Produsul Intern Brut (PIB) al României a reprezentat 90% din cel al anului precedent; sau că societatea comercială X a realizat în luna decembrie 2005 o valoare a vânzărilor de 3 ori mai mare decât valoarea programată pentru această lună. De asemenea, indicii preţurilor, inflaţia, s-au bucurat întotdeauna de interes, chiar şi din partea publicului larg, care a văzut în ei un mod de caracterizare, de oglindire a realităţii cotidiene, alături de indicele puterii de cumpărare sau indicele nivelului de trai.

În acest capitol vom vedea care este metodologia de calcul a indicilor ca măsură a variabilităţii fenomenelor, la nivel individual şi total, regulile de construire a indicilor de grup, cum se poate identifica şi măsura efectul influenţei factorilor ce acţionează asupra fenomenelor complexe.

Astăzi, indicii, datorită utilităţii lor, sunt aplicaţi într-o gamă lărgită de domenii ale vieţii economice, sociale, politice etc.

7.1. Mărimi relative. Conţinut, formă generală, tipologie

Mărimea relativă este un indicator statistic derivat, calculat sub formă de raport. La numărătorul fracţiei prin care se determină mărimea relativă întâlnim indicatorul raportat (comparat), iar la numitorul ei – indicatorul bază de comparaţie (bază de raportre). Forma generală a unei mărimi relative este:

$$Z = \frac{X}{Y} \cdot 10^k \qquad (7.1.)$$

unde:

Z este mărimea relativă;

X este indicatorul comparat;

Y este indicatorul bază de comparaţie;

k este o valoare întreagă care reglează forma de exprimare a mărimii relative. În funcţie de diferenţa dintre ordinul de mărime al indicatorului comparat şi al indicatorului bază de comparaţie, k poate lua următoarele valori:

Scanned with OKEN Scanner

- $k = 0$ dacă indicatorii de la numărătorul şi numitorul mărimii relative au aproximativ acelaşi ordin de mărime. În acest caz, mărimea relativă se exprimă în coeficienţi şi arată câte unităţi din indicatorul comparat revin pe o unitate a indicatorului bază de comparaţie;
- $k = 2$ dacă ordinul de mărime al indicatorului comparat este mai mic decât ordinul de mărime al indicatorului bază de comparaţie. Spunem că mărimea relativă este exprimată în procente şi arată câte unităţi din indicatorul comparat revin la o sută de unităţi ale indicatorului bază de comparaţie;
- $k = 3$ dacă indicatorul comparat este şi mai mic, ca ordin de mărime, decât indicatorul bază de comparaţie. Mărimea relativă va fi, în acest caz, exprimată în promile şi arată câte unităţi din indicatorul comparat revin la o mie de unităţi ale indicatorului bază de comparaţie;
- $k = 3$ sau $k = 4$ dacă numărătorul este mult mai mic decât numitorul, mărimea relativă fiind exprimată în prodecimile, respectiv procentimile.

X şi Y pot fi indicatori de acelaşi fel sau indicatori de natură diferită, care sunt comparabili, astfel încât mărimea relativă definită să aibă un conţinut real. Baza de raportare va fi constituită dintr-un indicator semnificativ, reprezentativ pentru evoluţia în profil temporal sau teritorial a indicatorului respectiv. Mărimile relative pot avea unităţi de măsură complexe, ce depind de unităţile de măsură ale celor doi indicatori comparaţi (dacă se compară indicatori de natură diferită) sau prin coeficienţi, respectiv procente, promile etc. (dacă se compară indicatori de acelaşi fel).

Principalele tipuri de mărimi relative sunt:
– mărimi relative de structură;
– mărimi relative de coordonare;
– mărimi relative de intensitate;
– mărimi relative de dinamică;
– mărimi relative ale programului.

A. Mărimile relative de structură

Se calculează sub forma: $\frac{parte}{întreg}$ ***, arătând modul de structurare al colectivităţii după variabila analizată.*** Mărimile relative de structură pot îmbrăca două forme:

– ***frecvenţe relative*** (pe care le-am studiat în capitolul 3), care se pot calcula cu una din relaţiile:

Scanned with OKEN Scanner

$$n_i^* = \frac{n_i}{\sum_{i=1}^{r} n_i} \qquad \text{(exprimate în coeficienți)} \qquad (7.2.)$$

sau:

$$n_{i\%}^* = \frac{n_i}{\sum_{i=1}^{r} n_i} \cdot 100 \qquad \text{(exprimate în procente)} \qquad (7.3.)$$

unde:

n_i este numărul de unități statistice dintr-o grupă i a colectivității ($i = \overline{1, r}$).

– ***greutăți specifice,*** care se pot calcula cu una din relațiile:

$$g_i^x = \frac{x_i}{\sum_{i=1}^{n} x_i} \qquad \text{(exprimate în coeficienți)} \qquad (7.4.)$$

sau:

$$g_{i\%}^x = \frac{x_i}{\sum_{i=1}^{n} x_i} \cdot 100 \qquad \text{(exprimate în procente)} \qquad (7.5.)$$

Greutățile specifice arată ponderea nivelului unei caracteristici X aferent unei unități sau grupe i (x_i) în nivelul total al caracteristicii ($\sum_{i=1}^{n} x_i$). Această relație este valabilă doar dacă valorile individuale ale variabilei X sunt însumabile.

Suma mărimilor relative de structură calculate pentru o aceeași colectivitate este 1 (dacă sunt exprimate în coeficienți) sau 100 (dacă sunt exprimate în procente):

$$\sum_{i=1}^{r} n_i^* = 1 \text{ sau } \sum_{i=1}^{r} n_{i\%}^* = 100 \qquad (7.6.)$$

$$\sum_{i=1}^{n} g_i^x = 1 \text{ sau } \sum_{i=1}^{n} g_{i\%}^x = 100 \qquad (7.7.)$$

Dacă populația este împărțită în grupe și s-au centralizat nivelurile variabilei pe aceste grupe și pe întreaga colectivitate, atunci greutatea specifică se va calcula după relația:

Scanned with OKEN Scanner

$$g_i^x = \frac{\sum_{j=1}^{ni} x_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{ni} x_{ij}} \quad \text{(exprimată în coeficienți)} \qquad (7.8)$$

sau

$$g_{i\%}^x = \frac{\sum_{j=1}^{ni} x_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{ni} x_{ij}} \cdot 100 \quad \text{(exprimată în procente)} \qquad (7.9)$$

unde:

n_i este numărul de unități statistice dintr-o grupă i a colectivității ($i = \overline{1, r}$);

x_{ij} este nivelul variabilei X pentru unitatea j din grupa i;

$\sum_{j=1}^{ni} x_{ij}$ reprezintă nivelul centralizat al variabilei X pentru grupa i;

$\sum_{i=1}^{r}\sum_{j=1}^{ni} x_{ij}$ reprezintă nivelul centralizat al variabilei X pe întreaga colectivitate.

Exemplul 7.1. Pentru 3 regiuni turistice se cunosc: numărul pensiunilor agroturistice şi numărul locurilor de cazare din cadrul acestora (Tabelul 7.1.).

Tabelul 7.1.

Regiunea turistică	Numărul pensiunilor agroturistice (ni)	Numărul locurilor de cazare (x_{ij})
A	7	20, 25, 12, 15, 20, 15, 18
B	10	40, 32, 12, 15, 25, 36, 30, 28, 20, 12
C	8	16, 14, 20, 25, 30, 24, 18, 16
Total	25	

Notând cu x_{ij} numărul locurilor de cazare din regiunea turistică i, pensiunea j, şi cu n_i numărul pensiunilor din regiunea i, putem determina următoarele mărimi relative de structură:

– pentru numărul locurilor de cazare:

- ponderea locurilor de cazare pe care le oferă o pensiune, în numărul total al locurilor de cazare din cadrul regiunii respective;

exemplu: $g_{A1}^x = \frac{20}{125} \cdot 100 = 16\%$, $g_{A2}^x = \frac{25}{125} \cdot 100 = 20\%$ etc.

Scanned with OKEN Scanner

- ponderea locurilor de cazare dintr-o regiune turistică în numărul total de locuri de cazare, oferite de toate cele 3 regiuni: $g_i^x = \frac{x_i}{\sum_{i=1}^{n} x_i}$, calculate în Tabelul 7.2., coloana 3.

Tabelul 7.2.

Regiunea turistică (0)	Numărul pensiunilor agroturistice (n_i) (1)	Nr. total al locurilor de cazare din cadrul regiunii (x_i) (2)	g_i^x (%) (3)	n_i^* (%) (4)	$k_{i/A}^x$ (5)	$k_{i/A}^n$ (6)	L_i (7)
A	7	125	23,2	28	1,0	1,0	17,86≈18
B	10	250	46,5	40	2,0	1,43	25
C	8	163	30,3	32	1.3	1,14	20,37≈20
Total	25	538	100	100			

Pentru a calcula aceste mărimi relative de structură, s-a determinat în prealabil, prin agregare, numărul total de locuri de cazare din fiecare regiune turistică:

$$\sum_{j=1}^{ni} x_{ij} = x_i \text{ (Tabelul 7.2., coloana 2)}$$

– pentru numărul de pensiuni agroturistice, se determină ponderea numărului de pensiuni dintr-o zonă în numărul total de pensiuni (adică frecvențele relative n_i^* – Tabelul 7.2., coloana 4).

Dacă datele au fost sistematizate după o variabilă numerică pe variante sau intervale de variație şi se cunosc frecvențele grupelor (n_i) şi variantele sau centrele de interval ale caracteristicii de grupare (x_i), atunci greutățile specifice se pot calcula cu relația:

$$g_i^x = \frac{x_i n_i}{\sum_{i=1}^{r} x_i n_i} \quad \text{(exprimate în coeficienți)} \qquad (7.10.)$$

sau

$$g_{i\%}^x = \frac{x_i n_i}{\sum_{i=1}^{r} x_i n_i} \cdot 100 \quad \text{(exprimate în procente)} \qquad (7.11.)$$

Mărimile relative de structură se reprezintă grafic prin diagrame de structură, care au fost prezentate pe larg în capitolul 2.

Scanned with OKEN Scanner

B. Mărimile relative de coordonare

Compară, prin prisma unui acelaşi indicator statistic, două unităţi statistice, două grupe ale aceleiaşi colectivităţi, sau două colectivităţi diferite şi se determină raportând nivelul indicatorului statistic pentru unitatea/ grupa/colectivitatea comparată cu nivelul aceluiaşi indicator statistic pentru unitatea/grupa/colectivitatea aleasă bază de comparaţie, obţinându-se un raport de coordonare:

$$k^{x}_{i/j} = \frac{x_i}{x_j} \qquad (7.12.)$$

unde:

$k^{x}_{i/j}$ este raportul de coordonare;

x_i este nivelul indicatorului X înregistrat la unitatea/grupa/colectivitatea i;

x_j este nivelul indicatorului X înregistrat la unitatea/grupa/colectivitatea j.

Raportul de coordonare ne arată de câte ori este mai mare sau mai mic (după cum este o mărime supraunitară sau subunitară) nivelul indicatorului în unitatea/grupa/colectivitatea comparată, faţă de nivelul indicatorului din unitatea/grupa/colectivitatea bază de comparaţie. Dacă vom compara în spaţiu două unităţi/grupe/colectivităţi diferite, atunci se va alege ca unitate bază de comparaţie o unitate teritorială mai semnificativă, unitatea teritorială cu nivelul cel mai scăzut/ridicat al indicatorului. Deseori, este permisă utilizarea unor baze reversibile, obţinându-se:

$$k^{x}_{j/i} = \frac{x_j}{x_i} \qquad (7.13.)$$

cu

$$k^{x}_{i/j} \cdot k^{x}_{j/i} = 1 \qquad (7.14.)$$

Raportul de coordonare se poate exprima şi procentual. Mărimile relative de coordonare se pot reprezenta grafic prin diagrame prin benzi, coloane sau suprafeţe.

Exemplul 7.2. Pentru problema din exemplul anterior, se pot calcula următoarele rapoarte de coordonare:

– pentru locurile de cazare turistică: se alege ca bază de comparaţie regiunea cu cel mai mic număr de locuri de cazare (A) şi se raportează numărul locurilor de cazare din regiunile B, respectiv C, la numărul locurilor de cazare din regiunea A, pentru a vedea de câte ori este mai mare numărul locurilor de cazare din regiunile B, C decât cel din A (Tabelul 7.2., coloana 5).

Scanned with OKEN Scanner

– pentru numărul de pensiuni: se alege tot regiunea A ca bază de comparație, deoarece are cele mai puține pensiuni, apoi se raportează numărul pensiunilor turistice din regiunile B, respectiv C, la numărul pensiunilor din regiunea A şi aflăm de câte ori sunt mai numeroase pensiunile din regiunile turistice B şi C decât cele din A. (Tabelul 7.2., coloana 6).

C. Mărimile relative de intensitate

Se calculează raportând doi indicatori de natură diferită, aferenți aceleiași unități/grupe/colectivități. Unitatea de măsură a mărimilor relative de intensitate este una complexă, rezultând din raportul unităților de măsură ale celor doi indicatori comparați. Indicatorii care se raportează trebuie să fie în *relație de interdependență*.

Pentru o unitate statistică *i*, formula de calcul a mărimii relative de intensitate este:

$$z_i = \frac{x_i}{y_i} \qquad (7.15.)$$

La nivelul întregii colectivități, mărimea relativă de intensitate se scrie ca raport între nivelurile agregate ale celor 2 indicatori raportați:

$$z = \frac{\sum x_i}{\sum y_i} \qquad (7.16.)$$

(dacă valorile individuale ale acestora sunt însumabile).

Dacă explicităm pe x_i din relația (7.15.) şi îl înlocuim în relația (7.16.), obținem:

$$x_i = z_i \cdot y_i \qquad (7.17.)$$

$$z = \frac{\sum z_i y_i}{\sum y_i} \qquad (7.18.)$$

Din relația (7.18.) rezultă că nivelul total al mărimii relative de intensitate se scrie ca medie aritmetică ponderată a mărimilor relative de intensitate la nivel individual, deci, la nivelul întregii populații, *mărimea relativă de intensitate are caracter de mărime medie*. Exemple de mărimi relative de intensitate sunt: productivitatea muncii, calculată ca raport între nivelul producției obținute şi numărul personalului productiv, salariul mediu – ca raport între fondul de salarii şi numărul de salariați etc. Mărimile relative de intensitate nu sunt însumabile şi se pot reprezenta grafic prin diagrame prin coloane, benzi sau suprafețe.

Scanned with OKEN Scanner

Exemplul 7.3. Pentru problema din exemplul (7.1.), se pot calcula următoarele mărimi relative de intensitate: numărul mediu de locuri de cazare al unei pensiuni, pe fiecare regiune şi pe total:

$$L_i = \frac{x_i}{n_i}$$

unde L_i este numărul mediu de locuri de cazare dintr-o pensiune, în regiunea turistică i (Tabelul 7.2., coloana 7).

Pe total, numărul mediu de locuri de cazare va fi:

$$\bar{L} = \frac{\sum x_i}{\sum n_i} = \frac{538}{25} = 21{,}52 \approx 22 \text{ locuri/pensiune}$$

D. Mărimile relative de dinamică

Compară, sub formă de raport, nivelurile realizate în două perioade sau momente de timp diferite ale aceluiaşi indicator. Rezultatul obţinut se numeşte *indice de dinamică* şi a fost studiat în capitolul 6. Indicele de dinamică, la nivelul unităţii/grupei i, se determină cu relaţia:

$$i^x_{1/0} = \frac{x^i_1}{x^i_0} \quad \text{sau} \quad i^x_{1/0} = \frac{x^i_1}{x^i_0} \cdot 100 \qquad (7.19.)$$

unde:

x^i_0 este nivelul realizat al indicatorului în perioada de bază pentru grupa i;

x^i_1 este nivelul realizat al indicatorului în perioada curentă pentru grupa i.

La nivelul întregii colectivităţi, indicele de dinamică se poate determina:

$$I^{\sum x}_{1/0} = \frac{\sum x^i_1}{\sum x^i_0} = \frac{\sum i^x_{1/0} \cdot x^i_0}{\sum x^i_0} = \frac{\sum x^i_1}{\sum \frac{x^i_1}{i^x_{1/0}}} \qquad (7.20.)$$

Mărimea relativă de dinamică se reprezintă grafic prin coloane, benzi sau suprafeţe.

E. Mărimile relative ale programului

În activitatea economico-socială, atât la nivel microeconomic, cât şi la nivel macroeconomic, este necesară programarea variabilelor ce caracterizează

Scanned with OKEN Scanner

activitatea desfăşurată, precum şi compararea nivelului realizat cu cel programat al variabilelor urmărite.

Se determină următoarele mărimi relative ale programului:

– ***mărimea relativă a sarcinii programate***, care se obţine prin raportarea nivelului programat al unei variabile în perioada curentă (x_p) la nivelul realizat al acesteia în perioada de bază (x_0), obţinându-se un indice:

$$i^x_{p/0} = \frac{x^i_p}{x^i_0} \qquad \text{(exprimat în coeficienţi)} \qquad (7.21.)$$

sau

$$i^x_{p/0} = \frac{x^i_p}{x^i_0} \cdot 100 \quad \text{(exprimat în procente)} \qquad (7.22.)$$

Indicii din relaţiile (7.21.) şi (7.22.) s-au calculat la nivel de unitate statistică (i). La nivel total, mărimea relativă a sarcinii programate se determină:

$$I^{\sum x}_{p/0} = \frac{\sum x^i_p}{\sum x^i_0} = \frac{\sum i^x_{p/0} \cdot x^i_0}{\sum x^i_0} \qquad (7.23.)$$

respectiv ca o medie aritmetică ponderată a indicilor individuali ai sarcinii programate.

– ***mărimea relativă a îndeplinirii programului***, care se obţine prin raportarea nivelului realizat al unei variabile în perioada curentă (x_1) la nivelul programat al acesteia în aceeaşi perioadă (x_p), obţinându-se, de asemenea, un indice, de asemenea la nivelul unităţii statistice i:

$$i^x_{1/p} = \frac{x^i_1}{x^i_p} \qquad \text{(exprimat în coeficienţi)} \qquad (7.24.)$$

sau

$$i^x_{1/p} = \frac{x^i_1}{x^i_p} \cdot 100 \quad \text{(exprimat în procente)} \qquad (7.25.)$$

La nivel total, mărimea relativă a îndeplinirii programului se determină ca o medie aritmetică ponderată a indicilor individuali ai îndeplinirii programului:

$$I^{\sum x}_{1/p} = \frac{\sum x^i_1}{\sum x^i_p} = \frac{\sum i^x_{1/p} \cdot x^i_p}{\sum x^i_p} \qquad (7.26.)$$

Mărimile relative ale programului se reprezintă grafic prin coloane, benzi sau suprafeţe. Între cei doi indici ai programului şi indicele de dinamică există următoarea relaţie: indicele de dinamică este egal cu produsul celor doi indici ai programului, conform relaţiei:

Scanned with OKEN Scanner

$$i^x_{1/0} = i^x_{1/p} \cdot i^x_{p/0} \quad \text{(la nivelul unei unități statistice)} \qquad (7.27.)$$

$$I^{\sum x}_{1/0} = I^{\sum x}_{1/p} \cdot I^{\sum x}_{p/0} \quad \text{(la nivelul întregii colectivități)} \qquad (7.28.)$$

7.2. Indicii ca mărimi relative. Rolul și funcțiile indicilor în cercetarea statistică

Indicii sunt o măsură a variației relative a fenomenelor, variație care poate fi observată și analizată după criterii de spațiu, timp și în funcție de un anumit sistem de referință. Fenomenele social-economice se manifestă după conținutul lor ca fenomene simple sau complexe, rezultate din acțiunea aditivă sau multiplicativă a mai multor factori.

Studiul mai aprofundat al variabilității, deși i-a preocupat pe oameni din cele mai vechi timpuri, a fost abordat într-o manieră științifică pe la mijlocul secolului al XVIII-lea.

Putem vorbi de noțiunea de *indice* la începutul secolului al XVIII-lea, într-un studiu al unui episcop anglican, privind evoluția prețurilor în Anglia între anii 1440 și 1700.

În cadrul unei definiții din 1970 din *Petit Robert*, se arată că indicele evidențiază, reflectă un fenomen, fără a da certitudinea că acest fenomen există.

Definiție:
Indicele statistic este o mărime relativă, care, ca procedeu de calcul, reprezintă un raport prin care se compară mărimea aceluiași fenomen înregistrat în două unități de timp, de spațiu, de program la o unitate statistică, la o grupă sau la nivelul întregii colectivități.

Numărătorul indicelui reprezintă volumul, nivelul fenomenului studiat în unitatea de timp/spațiu care se compară, iar numitorul acestuia – nivelul (volumul) fenomenului în unitatea de timp/spațiu aleasă ca bază de comparație (raportare). În alte cazuri, indicii pot compara nivelul realizat cu cel programat propus al fenomenului.

Nivelul care se constituie în bază de comparație, trebuie să fie un nivel reprezentativ, cu o semnificație deosebită în dezvoltarea fenomenului reflectat.

Deoarece indicii sunt raportarea a două niveluri diferite ale aceluiași fenomen, înseamnă că numărătorul și numitorul au aceeași unitate de măsură, și deci

Scanned with OKEN Scanner

indicele este adimensional, nu depinde de unitatea de măsură a fenomenului pentru care s-a calculat.

Indicele se exprimă în coeficienţi sau în procente. El arată de câte ori (de cât la sută) nivelul comparat al fenomenului este mai mare sau mai mic decât nivelul ales ca bază de comparaţie a fenomenului.

Uneori, indicii au şi un caracter de mărime medie, atunci când sunt calculaţi la nivelul întregului ansamblu sau al unei grupe a acestuia.

Pentru ca indicii calculaţi să aibă un conţinut real, se impune ca: datele iniţiale pe baza cărora s-a calculat indicele să fie complete, obţinute dintr-o observare totală (în cazul în care ele provin dintr-o observare parţială trebuie asigurată reprezentativitatea faţă de întreg); numărătorul şi numitorul să aibă valori reale cât mai posibile; diferenţa dintre numărătorul şi numitorul indicelui să fie egală cu modificarea absolută a fenomenului respectiv.

Deoarece ei se calculează, de regulă, pentru fenomene complexe, care sunt afectate de o gamă largă de factori de influenţă, indicii rezultaţi formează un sistem de indici (de exemplu, indicii producţiei, productivităţii muncii şi ai forţei de muncă formează un sistem de indici).

Metoda indicilor permite explicarea, descompunerea variaţiei fenomenelor complexe, ca urmare a modificării factorilor specifici folosind un întreg sistem de indici. Din acest punct de vedere, spunem că metoda indicilor, ca şi metoda corelaţiei, analizei dispersionale, este o metodă de analiză factorială.

7.3. Clasificarea indicilor

Datorită numărului mare şi varietăţii indicilor statistici, se impune o sistematizare a acestora, după următoarele criterii:

1) După **sfera de cuprindere a fenomenelor** (sau după nivelul la care se calculează), avem:

– indici individuali (elementari) – calculaţi la nivelul unei unităţi statistice;

– indici de grup (sintetici) – determinaţi la nivelul unei grupe a colectivităţii sau la nivelul întregii colectivităţi.

2) După **dimensiunea de raportare a fenomenului**:

– în raport cu timpul: indici cronologici (de dinamică);

– în raport cu spaţiul: indici teritoriali;

– în raport cu planul, programul: indici ai planului (ai îndeplinirii planului, ai sarcinii de plan).

3) După **natura variabilelor indexate**:

– indici ai variabilelor cantitative;

– indici ai variabilelor calitative.

Scanned with OKEN Scanner

4) După **modul de calcul**. Din punctul de vedere al acestui criteriu, indicii de grup pot fi:

– indici agregați;

– indici calculați ca medie a indicilor individuali;

– indici calculați ca raport a două medii.

5) După **natura ponderilor folosite**. Conform acestui criteriu, într-o serie de indici, indicii de grup pot fi:

– cu ponderi fixe (constante) – când se folosesc aceleaşi ponderi în întreaga serie;

– cu ponderi variabile – când ponderea folosită se schimbă odată cu schimbarea bazei de raportare.

6) După **baza de raportare** – indicii cronologici, după acest criteriu pot fi:

– cu bază fixă;

– cu bază mobilă sau în lanț.

7) După **natura fenomenului** pentru care se calculează indicii:

– indici ai valorii;

– indici ai volumului fizic;

– indici ai prețurilor;

– indici ai productivității muncii;

– indici ai salariului etc.

7.4. Reguli de construire a indicilor elementari şi sintetici

În continuare ne vom concentra asupra indicilor de dinamică, tratând modul de construire şi problemele ce pot apărea la calculul indicilor elementari şi sintetici.

7.4.1. Indici individuali (elementari)

Exprimă variația, sub formă de raport, a fenomenului la nivelul unei singure unități statistice.

Modul de calcul este simplu, precis, comparându-se prin raportare mărimea aceluiaşi fenomen înregistrat în două perioade de timp diferite. Una din aceste perioade este perioada comparată (de regulă se mai numeşte „perioada curentă" şi se notează cu indicativul *1*) şi apare în numărătorul fracției, iar cealaltă perioadă, cea de la numitor, este perioada de bază (sau perioada bază de comparație – cu indicativul *0*). Alegerea celor două perioade trebuie făcută cu precauție, astfel încât nivelurile fenomenului în cele două perioade să fie comparabile.

Scanned with OKEN Scanner

Fie y o variabilă complexă, pentru care indicele calculat va fi:

$$i_{1/0}^{y} = \frac{y_1^i}{y_0^i} \text{ sau } i_{1/0}^{y} = \frac{y_1^i}{y_0^i} \cdot 100 \qquad (7.29.)$$

Valoarea indicelui poate fi egală sau diferită de 1.

• dacă $i_{1/0}^{y} = 1$, avem o stagnare a fenomenului, adică $y_1^i = y_0^i$, la nivelul unității i;

• dacă $i_{1/0}^{y} \neq 1$, înseamnă că s-a produs o modificare astfel:

– $i_{1/0}^{y} > 1 \Rightarrow$ a avut loc o creştere ($y_1^i > y_0^i$)

– $i_{1/0}^{y} < 1 \Rightarrow$ a avut loc o scădere ($y_1^i < y_0^i$)

Dacă, aşa cum am arătat, y este o variabilă complexă, ea se poate descompune într-un produs de factori. Aceşti factori se pot împărţi în două categorii mari:

- *factori cantitativi,* de natură extensivă, care se notează cu f_i (în această categorie se poate înscrie şi numărul de unităţi statistice, confundându-se în acest caz cu frecvenţele de apariţie);
- *factori calitativi,* de natură intensivă, simbolizaţi cu x_i (pot apărea sub forma valorilor caracteristicilor luate în studiu).

La nivelul unei unităţi statistice i, nivelul caracteristicii complexe y_i se poate scrie ca produs al celor două componente (cantitativă şi calitativă) la nivelul unităţii i.

$$y_i = x_i f_i, i = \overline{1,n} \qquad (7.30.)$$

Pentru cele două variabile factoriale (x şi f) se pot scrie doi indici individuali conform relaţiilor:

$$i_{1/0}^{x} = \frac{x_1^i}{x_0^i}, i = \overline{1,n} \qquad (7.31.)$$

$$i_{1/0}^{f} = \frac{f_1^i}{f_0^i}, i = \overline{1,n} \qquad (7.32.)$$

Indicele individual al variabilei complexe y se mai scrie:

$$i_{0/0}^{y} = \frac{y_1^i}{y_0^i} = \frac{x_1^i f_1^i}{x_0^i f_0^i} = \frac{x_1^i}{x_0^i} \cdot \frac{f_1^i}{f_0^i} = i_{1/0}^{x} \cdot i_{1/0}^{f} \qquad (7.33.)$$

Aşadar, s-a pus în evidenţă o relaţie între cei trei indicatori, şi anume: indicele individual al variabilei complexe este egal cu produsul indicilor individuali ai celor doi factori.

Scanned with OKEN Scanner

7.4.2. Indici de grup (sintetici)

Reflectă variația medie relativă la nivelul întregii colectivități sau al unei grupe a acesteia. Se notează de regulă cu litere mari (I).

Înainte de construirea indicilor de grup, trebuie efectuată o analiză complexă a fenomenului studiat și trebuie răspuns la următoarele probleme: cum pot fi separați corect factorii cantitativi de cei calitativi; dacă acești factori sunt direct însumabili sau nu; dacă nu – ce sistem de pondere trebuie folosit; ce relație de calcul trebuie folosită pentru calculul indicelui de grup etc.

După modul de calcul, indicii de grup se împart în trei categorii:

a) indici agregați;

b) indici calculați ca medie a indicilor individuali;

c) indici calculați ca raport a două medii.

7.4.2.1. Indici sintetici calculați ca indici agregați

La nivelul întregii colectivități, volumul total al variabilei complexe se obține prin însumarea (agregarea) valorilor înregistrate la nivelul fiecărei unități statistice. Prin raportarea nivelului agregat al fenomenului din perioada comparată (curentă) ($\sum y_1$) la cel din perioada de bază ($\sum y_0$) se obține indicele agregat.

$$\Sigma y^i = \Sigma x^i f^i \rightarrow \Sigma y_1^i = \Sigma x_1^i f_1^i \qquad (7.34.)$$

$$\rightarrow \Sigma y_0^i = \Sigma x_0^i f_0^i \qquad (7.35.)$$

Indicele de grup al caracteristicii complexe y este:

$$I_{1/0}^{y} = \frac{\sum_i y_1^i}{\sum_i y_0^i} = \frac{\sum_i x_1^i f_1^i}{\sum_i x_0^i f_0^i} \qquad (7.36.)$$

De cele mai multe ori, elementele din care se compun fenomenele din natură și societate sunt eterogene, de aceea ele nu pot fi însumate direct. Pentru a face însumabile aceste elemente se apelează la un comăsurător, etalon, numit și *pondere*.

Ponderea este întâlnită atât în numărătorul, cât și numitorul indicelui de grup, cu aceeași valoare.

Pentru a determina nivelul agregat al valorilor celor doi factori identificați ai variabilei complexe (factorul cantitativ și calitativ) se pune problema însumabilității valorilor lor individuale.

În cazul valorilor factorului cantitativ pot exista două situații. Dacă este vorba de unități fizice, naturale de același fel, atunci însumarea lor se face în mod direct (de exemplu, personalul angajat în trei filiale ale unei societăți, comerciale) (Σf^i). De cele mai multe ori, însă, valorile individuale ale factorului extensiv nu

Scanned with OKEN Scanner

pot fi însumabile, întrucât s-ar agrega unităţi fizice de natură diferită (de exemplu, cantităţi din 3 produse diferite). În acest ultim caz se înmulţesc valorile factorului extensiv cu ponderile. La fel se întâmplă şi în cazul factorului calitativ (intensiv), care în majoritatea cazurilor are valori neaditive.

Rolul de pondere îl poate juca factorul pereche al factorului ce trebuie agregat. Ponderea poate fi reprezentată, deci, atât de valorile individuale ale factorului cantitativ, cât şi ale celui calitativ.

Pe lângă indicele de grup al variabilei complexe se calculează şi doi indici de grup ai celor doi factori de influenţă (numiţi şi indici factoriali).

Indicele factorial al factorului cantitativ se scrie:

$$I^{f}_{1/0} = \frac{\sum_{i} x^{i} f^{i}_{1}}{\sum_{i} x^{i} f^{i}_{0}}, \text{ unde } \quad x = \text{pondere}$$

i = unitatea statistică (7.37.)

Indicele factorial al factorului calitativ se calculează conform relaţiei:

$$I^{x}_{1/0} = \frac{\sum_{i} x^{i}_{1} f^{i}}{\sum_{i} x^{i}_{0} f^{i}}, \text{ unde } \quad f = \text{pondere}$$

i = unitatea statistică (7.38.)

Ponderile pot fi la nivelul perioadei de bază sau la nivelul perioadei curente. Alegerea acestei perioade se concretizează, de fapt, în alegerea sistemului de ponderare adecvat. În continuare, pentru uşurinţă, se va renunţa la notaţia i a unităţii statistice.

Indicele factorial ce arată influenţa factorului cantitativ (extensiv) va avea una din formele:

– dacă ponderile sunt la nivelul perioadei de bază:

$$I^{f}_{1/0} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \qquad (7.39.)$$

– dacă ponderile sunt la nivelul perioadei curente:

$$I^{f}_{1/0} = \frac{\Sigma x_1 f_1}{\Sigma x_1 f_0} \qquad (7.40.)$$

Analog, indicele factorial ce arată influenţa variabilei calitative (intensive) va fi:

– dacă ponderile sunt la nivelul perioadei de bază:

$$I^{x}_{1/0} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \qquad (7.41.)$$

Scanned with OKEN Scanner

– dacă ponderile sunt la nivelul perioadei curente:

$$I_{1/0}^{x} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \quad (7.42.)$$

Produsul indicilor factoriali este egal cu indicele variabilei complexe (*testul de reversibilitate al factorilor*) doar dacă în construirea indicilor factoriali se vor utiliza sisteme de ponderare diferite.

Modificările absolute se vor constitui pornind de la formulele de calcul ale indicilor, şi anume scăzând numitorul din numărător.

$$\Delta\Sigma y = \Sigma y_1 - \Sigma y_0 = \Sigma x_1 f_1 - \Sigma x_0 f_0 \quad (7.43.)$$

$$\Delta\Sigma y(f) \rightarrow \Delta\Sigma y(f) = \Sigma x_0 f_1 - \Sigma x_0 f_0 \text{ (pentru ponderi din perioada de bază)} \quad (7.44.)$$

$$\rightarrow \Delta\Sigma y(f) = \Sigma x_1 f_1 - \Sigma x_1 f_0 \text{ (pentru ponderi din perioada curentă)} \quad (7.45.)$$

$$\Delta\Sigma y(x) \rightarrow \Delta\Sigma y(x) = \Sigma x_1 f_0 - \Sigma x_0 f_0 \text{ (pentru ponderi din perioada de bază)} \quad (7.46.)$$

$$\rightarrow \Delta\Sigma y(x) = \Sigma x_1 f_1 - \Sigma x_0 f_1 \text{ (pentru ponderi din perioada curentă)} \quad (7.47.)$$

Suma modificărilor absolute ale factorilor este egală cu modificarea absolută a variabilei complexe dacă în construirea modificării absolute a unui factor se folosesc ponderi dintr-o perioadă, iar la construirea modificării absolute a celuilalt factor – ponderi din cealaltă perioadă.

Exemplul 7.4. Pentru indicele şi modificarea absolută a factorului cantitativ folosim ponderi din perioada de bază, iar pentru indicele şi modificarea absolută a factorului calitativ folosim ponderi din perioada curentă.

$$I_{1/0}^{x} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \quad (7.48.)$$

$$\Delta\Sigma y(x) = \Sigma x_1 f_1 - \Sigma x_0 f_1 \quad (7.49.)$$

$$I_{1/0}^{f} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \quad (7.50.)$$

$$\Delta\Sigma y(f) = \Sigma x_0 f_1 - \Sigma x_0 f_0 \quad (7.51.)$$

$$I_{1/0}^{\Sigma y} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_0} \quad (7.52.)$$

$$\Delta\Sigma y = \Sigma x_1 f_1 - \Sigma x_0 f_0 \quad (7.53.)$$

$$I_{1/0}^{\Sigma y} = I_{1/0}^{x} = I_{1/0}^{f} \quad (7.54.)$$

$$\Delta\Sigma y = \Delta\Sigma y(x) + \Delta\Sigma y(f) \quad (7.55.)$$

Scanned with OKEN Scanner

Exemplul 7.5. Despre cele trei mărfuri vândute la un magazin în noiembrie şi decembrie 2005 se cunosc datele (coloanele 0-4, Tabelul 7.3.):

Tabelul 7.3.

Marfa	Cantităţi vândute		Preţuri unitare (mii lei)		Valoarea vânzărilor (mii lei)	
	nov.'05 (q_0)	dec.'05 (q_1)	nov.'05 (p_0)	dec.'05 (p_1)	nov.'05 (v_0)	dec.'05 (v_1)
0	1	2	3	4	5	6
M1 (buc)	50	60	25	20	1 250	1 200
M2 (l)	20	50	10	7	200	350
M3 (kg)	35	5	8	30	280	150
Total	–	–	–	–	1 730	1 700

Se fac următoarele notaţii:

q_0, q_1 = cantităţile vândute din fiecare marfă în perioada de bază şi curentă;

p_0, p_1 = preţurile unitare ale fiecărei mărfi în cele două perioade;

v_0, v_1 = valoarea vânzărilor pentru fiecare marfă în cele 2 perioade.

Între cele 3 variabile, există relaţia:

$v = p\,q$

$v_0 = p_0 q_0$ (coloana 5, Tabelul 7.3.)

$v_1 = p_1 q_1$ (coloana 6, Tabelul 7.3.)

La nivel total, avem:

$$\Sigma v_0 = \Sigma p_0 q_0$$

$$\Sigma v_1 = \Sigma p_1 q_1$$

Indicele variaţiei valorii vânzărilor totale va fi:

$$I^{\Sigma v} = \frac{\Sigma v_1}{\Sigma v_0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{1700}{1730} = 0{,}98$$

Acesta se descompune în cei doi indici factoriali:

$$I^{p} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\Sigma v_1}{\Sigma p_0 q_1} = \frac{1700}{60 \cdot 25 + 50 \cdot 10 + 5 \cdot 8} = \frac{1700}{1500 + 500 + 40} = \frac{1700}{2040} = 0{,}83$$

$$I^{q} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_0 q_1}{\Sigma v_0} = \frac{2040}{1730} = 1{,}18$$

$$I^{\Sigma v} = I^{p} \cdot I^{q}$$

Modificările absolute vor fi:

$$\Delta\Sigma v = \Sigma v_1 - \Sigma v_0 = 1700 - 1730 = -30 \text{ mii lei}$$

$$\Delta\Sigma v(p) = \Sigma p_1 q_1 - \Sigma p_0 q_1 = 1700 - 2040 = -340 \text{ mii lei}$$

$$\Delta\Sigma v(q) = \Sigma p_0 q_1 - \Sigma p_0 q_0 = 2040 - 1730 = 310 \text{ mii lei}$$

cu $\Delta\Sigma v = \Delta\Sigma v(p) + \Delta\Sigma v(q)$

Scanned with OKEN Scanner

Rezultatele obținute arată că valoarea totală a mărfurilor vândute a scăzut în decembrie 2005 cu 2% față de luna anterioară, adică cu 30 mii lei. Dacă s-ar fi modificat numai prețurile unitare ale mărfurilor, valoarea vânzărilor ar fi scăzut cu 17% (adică cu 340 mii lei), iar dacă s-ar fi modificat doar cantitățile vândute, prețurile rămânând neschimbate, valoarea vânzărilor totale ar fi crescut de 1,18 ori (sau cu 18%), adică cu 310 mii lei în decembrie față de noiembrie 2005.

7.4.2.2. Indici sintetici calculați ca medie a indicilor individuali

Această metodă se aplică în calculul indicilor de grup ori de câte ori nu există suficiente informații pentru calculul indicilor agregați. Oricum, dacă ar exista posibilitatea aplicării ambelor metode, ar trebui ca valoarea indicelui de grup calculat ca medie a indicilor individuali să fie egală cu valoarea indicelui de grup calculat sub formă agregată.

Indicii de grup se pot forma fie ca medie aritmetică ponderată, fie ca medie armonică ponderată a indicilor individuali, în funcție de datele inițiale cunoscute.

• ca medie aritmetică (ponderată):

– pentru variabila cantitativă aditivă

$$I_{1/0}^{f} = \frac{\Sigma i^{f} \cdot f_0}{\Sigma f_0} \qquad (7.56.)$$

– pentru variabila cantitativă non-aditivă

$$I_{1/0}^{f} = \frac{\Sigma i^{f} x_0 f_0}{\Sigma x_0 f_0} = \frac{\Sigma i^{f} \cdot y_0}{\Sigma y_0} \qquad (7.57.)$$

• ca medie armonică (ponderată):

– pentru variabila cantitativă – aditivă

$$I_{1/0}^{f} = \frac{\Sigma f_1}{\Sigma \frac{1}{i^{f}} f_1} \qquad (7.58.)$$

– pentru variabila cantitativă – non-aditivă

$$I_{1/0}^{f} = \frac{\Sigma x_1 f_1}{\Sigma \frac{1}{i^{f}} x_1 f_1} = \frac{\Sigma y_1}{\Sigma \frac{1}{i^{f}} y_1} \qquad (7.59.)$$

– pentru variabila calitativă – (non-aditivă):

$$I_{1/0}^{x} = \frac{\Sigma x_1 f_1}{\Sigma \frac{1}{i^{x}} x_1 f_1} = \frac{\Sigma y_1}{\Sigma \frac{1}{i^{x}} y_1} \qquad (7.60.)$$

Scanned with OKEN Scanner

Aşadar, media aritmetică se foloseşte pentru calculul indicilor de grup ai variabilei cantitative, atunci când se cunosc indicii individuali ai acestei variabile $\left(i^f\right)$ şi nivelul total, agregat, al variabilei complexe din perioada de bază $\left(\Sigma y_0 = \Sigma x_0 f_0\right)$.

Media armonică se aplică la construirea ambilor indici de grup factoriali, atunci când se cunosc: indicii individuali ai factorului respectiv $\left(i^x, i^f\right)$ şi nivelul agregat totalizat, al variabilei complexe în perioada curentă $\left(\Sigma y_1\right)$.

De regulă, indicele de grup al factorului cantitativ se calculează ca medie aritmetică, ponderată, iar indicele de grup al factorului calitativ – ca medie armonică ponderală a indicilor individuali ai factorului respectiv.

Exemplul 7.6. Se cunosc următoarele date, privind vânzările a 3 produse (coloanele 0-3, Tabelul 7.4.):

Tabelul 7.4.

Produse	U.M.	Valoarea vânzărilor în perioada curentă v_1 (mil. lei)	Modificarea relativă a preţurilor unitare r^p (%)	i^p (%)
0	1	2	3	4
A	buc.	20	+ 10	110
B	litri	35	+ 5	105
C	kg	40	– 2	98
Total	–	95	–	–

Notăm cu:

q_0, q_1 – cantităţile vândute dintr-un produs în perioada de bază şi curentă;

p_0, p_1 – preţurile unitare de vânzare ale fiecărui produs în perioada de bază şi curentă;

v_0, v_1 – valoarea vânzărilor fiecărui produs în perioada de bază, respectiv curentă;

r^p – ritmul de modificare a preţurilor unitare (modificarea relativă).

Relaţia dintre indice şi ritm este:

$r_{\%}^p = \left(i^p - 1\right)100 \Rightarrow i_{\%}^p = r_{\%}^p + 100$ (coloana 4, Tabelul 7.2.)

Indicele factorial al preţului unitar se scrie:

$$I_{1/0}^p = \frac{\Sigma p_1 q_1}{\Sigma \frac{1}{i^p} p_1 q_1} = \frac{\Sigma v_1}{\Sigma \frac{1}{i^p} v_1} = \frac{95}{\frac{1}{1,1} \cdot 20 + \frac{1}{1,05} \cdot 35 + \frac{1}{0,98} \cdot 40} =$$

$$= \frac{95}{18,18 + 33,33 + 40,82} = \frac{95}{92,326} = 1,029 \approx 1,03$$

Scanned with OKEN Scanner

Corespunzător se poate calcula şi modificarea absolută a valorii vânzărilor datorate modificărilor preţurilor unitare:

$$\Delta v(p) = \Sigma p_1 q_1 - \Sigma \frac{1}{i^p} p_1 q_1 = 95 - 92{,}326 = 2{,}674 \text{ mil. lei}$$

Rezultatele obţinute arată că valoarea vânzărilor a crescut în perioada curentă faţă de cea de bază cu 3%, datorită modificării preţurilor unitare, adică, cu 2 674 milioane lei.

Exemplul 7.7. Despre producerea a trei mărfuri se cunosc datele (coloanele 0-3, Tabelul 7.5.):

Tabelul 7.5.

Marfă	Valoarea producţiei în per. de bază (mil. lei) (v_0)	Indice preţuri (%) (i^p)	Indicele valorii producţiei (i^v)	i^q
0	1	2	3	4
M1	15	120	1,02	0,85
M2	20	90	1,1	1,22
M3	40	105	1,5	1,43

Se folosesc notaţiile de la problema precedentă. Indicele de grup al valorii producţiei:

$$I^v_{1/0} = \frac{\Sigma v_1}{\Sigma v_0}$$

$$\text{Dar } \frac{v_1}{v_0} = i^v \Rightarrow v_1 = i^v \cdot v_0$$

$$I^v_{1/0} = \frac{\Sigma i^v v_0}{\Sigma v_0} = \frac{1{,}02 \cdot 15 + 1{,}1 \cdot 20 + 1{,}5 \cdot 40}{15 + 20 + 40} = \frac{15{,}3 + 22 + 60}{75} = \frac{97{,}3}{75} =$$

$$= 1{,}297 \approx 1{,}3\,(130\%)$$

Modificarea absolută a valorii producţiei va fi:

$$\Delta v = \Sigma v_1 - \Sigma v_0 = \Sigma i^v v_0 - \Sigma v_0 = 97{,}3 - 75 = 22{,}3 \text{ milioane lei}$$

Indicele de grup al volumului fizic:

$$I^q_{1/0} = \frac{\Sigma i^q q_0 p_0}{\Sigma q_0 p_0} = \frac{\Sigma i^q v_0}{\Sigma v_0}$$

$$i^v = i^p \cdot i^q \Rightarrow i^q = i^v / i^p \text{ (coloana 4, Tabelul 7.5.)}$$

$$I^q_{1/0} = \frac{0{,}85 \cdot 15 + 1{,}22 \cdot 20 + 1{,}43 \cdot 40}{75} = \frac{12{,}75 + 24{,}4 + 57{,}2}{75} =$$

$$= \frac{94{,}35}{75} = 1{,}258$$

Scanned with OKEN Scanner

Modificarea absolută a valorii producției ca urmare a variației cantităților vândute va fi:

$$\Delta q = \Sigma i^q \cdot q_0 p_0 - \Sigma q_0 p_0 = 94{,}35 - 75 = 19{,}35 \text{ milioane lei}$$

Indicele de grup al prețurilor:

$$I_{1/0}^{p} = \frac{\Sigma p_1 q_1}{\Sigma \frac{1}{i^p} p_1 q_1} = \frac{\Sigma v_1}{\Sigma \frac{1}{i^p} v_1} = \frac{97{,}3}{15{,}3 \cdot \frac{1}{1{,}2} + 22 \cdot \frac{1}{0{,}9} + 60 \cdot \frac{1}{1{,}05}} =$$

$$= \frac{97{,}3}{12{,}75 + 24{,}44 + 57{,}14} = \frac{97{,}3}{94{,}33} = 1{,}031$$

Modificarea absolută a valorii producției cauzată de variația prețurilor unitare va fi:

$$\Delta v = \Sigma v_1 - \Sigma v_0 = \Sigma p_1 q_1 - \Sigma \frac{1}{i^p} p_1 q_1 = 97{,}3 - 94{,}33 = 3 \text{ mil lei}$$

Valoarea totală a producției a crescut de 1,3 ori (cu 30%) sau cu 22,3 mil. lei. Sub influența modificării cantităților vândute, valoarea producției a crescut de 1,25 ori (sau cu 19,35 mil. lei), iar sub influența modificării prețului unitar de 1,03 ori (sau cu 3 mil. lei).

7.4.2.3. Indici de grup calculați ca raport a două medii

Această metodă de calcul se utilizează ori de câte ori se pune problema caracterizării variației unei caracteristici, a cărei valoare se formează la nivel de grupă sau de colectivitate totală, sub formă de medie. Acești indici se folosesc, de regulă, la caracteristicile calitative, în condițiile în care valorile individuale ale factorului cantitativ sunt însumabile (Σf):

$$y = x \cdot f \Rightarrow x = \frac{y}{f} \tag{7.61.}$$

$$\Sigma y = \bar{x} \cdot \Sigma f$$

$$\bar{x} = \frac{\Sigma y}{\Sigma f} = \frac{\Sigma xf}{\Sigma f} \rightarrow \bar{x}_1 = \frac{\Sigma y_1}{\Sigma f_1} = \frac{\Sigma x_1 f_1}{\Sigma f_1} \tag{7.62.}$$

$$\rightarrow \bar{x}_0 = \frac{\Sigma y_0}{\Sigma f_0} = \frac{\Sigma x_0 f_0}{\Sigma f_0} \tag{7.63.}$$

Valorile individuale ale caracteristicii calitative se obțin prin raportarea valorilor individuale ale variabilei complexe la cele ale variabilei cantitative, adică au caracter de mărime relativă de intensitate. La nivel de grup, însă, valoarea sintetică (totală) a variabilei calitative are caracter de mărime medie (se determină

Scanned with OKEN Scanner

ca medie aritmetică ponderată). Atunci şi indicele de variaţie a acestei caracteristici se calculează ca raport a două medii:

$$I_{SV}^{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0} \qquad (7.64.)$$

Variaţia fenomenului calitativ depinde, pe de o parte, de modificarea valorilor sale individuale, iar pe de altă parte, de modificarea frecvenţelor de apariţie a acestor valori în cele două perioade.

Dacă notăm: $f_i^* = \frac{f_i}{\Sigma f_i} \cdot 100 \rightarrow f_i^*$ sunt frecvenţe relative, atunci indicele de grup devine:

$$I_{1/0}^{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0} = \frac{\Sigma x_1 f_1^*}{\Sigma x_0 f_0^*} \qquad (7.65.)$$

Astfel, indicele calculat ca raport a două medii devine indice agregat. Variabila cantitativă influenţează modificarea mediei factorului calitativ şi prin modificarea structurii, a ponderii.

În funcţie de variaţia structurii frecvenţelor de apariţie ale valorilor individuale din care s-au calculat cele două medii, indicii calculaţi ca raport de medii se clasifică în:

- *indicele cu structură variabilă* – când modificarea nivelului mediu e influenţată atât de variaţia factorului calitativ, cât şi de variaţia celui cantitativ – relaţiile (7.64.) şi (7.65.).
- *indicele cu structură fixă* – prin care se măsoară influenţa modificării valorilor individuale ale variabilei calitative, în ipoteza menţinerii neschimbate a frecvenţelor lor de apariţie.

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f}{\Sigma f} : \frac{\Sigma x_0 f}{\Sigma f} = \frac{\Sigma x_1 f^*}{\Sigma x_0 f^*} \qquad (7.66.)$$

Frecvenţele pot fi luate din perioada curentă:

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_1}{\Sigma f_1} = \frac{\Sigma x_1 f_1^*}{\Sigma x_0 f_1^*} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} = \frac{\Sigma x_1 f_1}{\Sigma \frac{1}{i_{1/0}^{x}} x_1 f_1} \qquad (7.67.)$$

sau din perioada de bază:

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f_0}{\Sigma f_0} : \frac{\Sigma x_0 f_0}{\Sigma f_0} = \frac{\Sigma x_1 f_0^*}{\Sigma x_0 f_0^*} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} = \frac{\Sigma i_{1/0}^{f} x_0 f_0}{\Sigma x_0 f_0} \qquad (7.68.)$$

Acest indice se poate calcula şi ca medie aritmetică sau armonică ponderată.

Scanned with OKEN Scanner

• *indicele variației structurii* – se folosește pentru a caracteriza variația unui fenomen complex sub influența modificării numai a structurii factorului cantitativ, în condițiile menținerii neschimbate a variabilei calitative:

$$I^{\bar{x}(f)} = \frac{\Sigma x f_1}{\Sigma f_1} : \frac{\Sigma x f_0}{\Sigma f_0} \qquad (7.69.)$$

x poate lua valori:

– în perioada curentă:

$$I_{1/0}^{\bar{x}(f)} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_1 f_0}{\Sigma f_0} \qquad (7.70.)$$

– în perioada de bază:

$$I_{1/0}^{\bar{x}(f)} = \frac{\Sigma x_0 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0} \qquad (7.71.)$$

Indicele cu structură fixă și indicele variației structurii sunt indici factoriali ai indicelui cu structură variabilă, iar produsul indicilor factoriali este egal cu indicele cu structură variabilă:

$$I_{1/0}^{\bar{x}} = I_{1/0}^{\bar{x}(x)} \cdot I_{1/0}^{\bar{x}(f)} \qquad (7.72.)$$

Relațiile de mai sus se verifică numai dacă la indicii factoriali s-au folosit ponderi din perioade diferite (se obișnuiește, astfel, în practică ca la indicele cu structură fixă să se folosească ponderi din perioada curentă, iar la indicele variației structurii – ponderi din perioada de bază).

Modificările absolute ale mediei sub influența celor doi factori se scriu ca diferență între numărătorul și numitorul indicilor în cauză:

$$\Delta\bar{x} = \frac{\Sigma \bar{x}_1 f_1}{\Sigma f_1} - \frac{\Sigma \bar{x}_0 f_0}{\Sigma f_0} = \Sigma \bar{x}_1 f_1^* - \Sigma \bar{x}_0 f_0^* \qquad (7.73.)$$

$$\Delta\bar{x}(x) = \frac{\Sigma x_1 f_1}{\Sigma f_1} - \frac{\Sigma x_0 f_1}{\Sigma f_1} = \Sigma x_1 f_1^* - \Sigma x_0 f_1^* \qquad (7.74.)$$

$$\Delta\bar{x}(f) = \frac{\Sigma x_0 f_1}{\Sigma f_1} - \frac{\Sigma x_0 f_0}{\Sigma f_0} = \Sigma x_0 f_1^* - \Sigma x_0 f_0^* \qquad (7.75.)$$

Modificarea absolută a mediei este egală cu suma modificărilor absolute datorate factorilor de influență.

$$\Delta\bar{x}(x) + \Delta x(f) = \Delta\bar{x} \qquad (7.76.)$$

Exemplul 7.8. Pentru două filiale ale unei societăți comerciale, se cunosc fondul se salarii și numărul de salariați:

Scanned with OKEN Scanner

Tabelul 7.6.

Filiala	Fond salarii (mil.lei)		Nr. salariaţi (pers.)	
	Per. de bază (FS_0)	Per. curentă (FS_1)	Per. de bază (N_0)	Per. curentă (N_1)
0	1	2	3	4
A	40	63	20	18
B	75	140	30	35
Total	115	203	50	53

Vom face următoarele notaţii:

FS_0, FS_1 = fondul de salarii la nivel de filială, în perioada de bază şi curentă;

N_0, N_1 = nr. de salariaţi la nivel de filială, în perioada de bază şi curentă;

S_0 , S_1 = salariul mediu la nivelul unei filiale, în perioada de bază şi curentă;

N_0^*, N_1^* = structura salariaţilor în perioada de bază şi curentă.

$$S_0 = \frac{FS_0}{N_0};\ S_1 = \frac{FS_1}{N_1}$$

$$N_0^* = \frac{N_0}{\Sigma N_0};\ N_1^* = \frac{N_1}{\Sigma N_1}$$

Salariul mediu la nivelul întregii societăţi comerciale:

$$\bar{S}_0 = \frac{\Sigma FS_0}{\Sigma N_0} = \frac{115}{50} = 2{,}3 \text{ mil lei/pers.}$$

$$\bar{S}_1 = \frac{\Sigma FS_1}{\Sigma N_1} = \frac{203}{53} = 3{,}83 \text{ mil lei/pers.}$$

Salariul mediu pe total S.C. se mai poate determina şi ca medie aritmetică ponderată a salariilor individuale (de la nivelul filialelor):

$$\bar{S}_0 = \frac{\Sigma S_0 N_0}{\Sigma N_0} = \Sigma S_0 N_0^* = 2 \cdot 0{,}4 + 2{,}5 \cdot 0{,}6 = 0{,}8 + 1{,}5 = 2{,}3 \text{ mil lei/pers.}$$

$$\bar{S}_1 = \frac{\Sigma S_1 N_1}{\Sigma N_1} = \Sigma S_1 N_1^* = 3{,}5 \cdot 0{,}34 + 4 \cdot 0{,}66 = 1{,}19 + 2{,}64 = 3{,}83 \text{ mil lei/pers.}$$

Precizăm că salariul mediu la nivelul filialelor (S_0, S_1) şi structura salariaţilor (N_0^*, N_1^*) s-au calculat în Tabelul 7.7.

Scanned with OKEN Scanner

Tabelul 7.7.

Filiala	Salariul mediu (mil.lei/pers.)		Structura salariaților	
	Per. de bază	Per. curentă	Per. de bază	Per. curentă
0	1	2	3	4
A	2,0	3,5	0,4	0,34
B	2,5	4,0	0,6	0,66
Total	–	–	1,0	1,0

Indicele salariului mediu pe total va fi:

$$I^{\bar{S}} = \frac{\bar{S}_1}{\bar{S}_0} = \frac{3,83}{2,3} = 1,665\,(166,5\%)$$

care arată că salariul mediu total a crescut în perioada curentă față de perioada de bază de 1,665 ori (sau cu 66, 5 %).

În valoare absolută, această creştere a fost de

$$\Delta\bar{S} = \bar{S}_1 - \bar{S}_0 = 3,83 - 2,3 = 1,53 \text{ mii lei}$$

Pentru a vedea în ce măsură se datorează această variație celor doi factori de influență ai salariului mediu total (adică salariile la nivelul fiecărei filiale şi structura salariaților), vom calcula cei doi indici factoriali:

$$I^{\bar{S}(S)} = \frac{\Sigma S_1 N_1}{\Sigma N_1} : \frac{\Sigma S_0 N_1}{\Sigma N_1} = \frac{\Sigma S_1 N_1^*}{\Sigma S_0 N_1^*} = \frac{\bar{S}}{\Sigma S_0 N_1^*} = \frac{3,83}{2 \cdot 0,34 + 2,5 \cdot 0,66} =$$

$$= \frac{3,83}{0,68 + 1,65} = \frac{3,83}{2,33} = 1,643$$

$$I^{\bar{S}(N^*)} = \frac{\Sigma S_0 N_1}{\Sigma N_1} : \frac{\Sigma S_0 N_0}{\Sigma N_0} = \frac{\Sigma S_0 N_1^*}{\Sigma S_0 N_0^*} = \frac{\Sigma S_0 N_1^*}{\bar{S}_0} = \frac{2,33}{2,3} = 1,013$$

Se verifică, de asemenea, şi relația dintre indicele general şi indicii factoriali:

$$I^{\bar{S}(S,N^*)} = I^{\bar{S}(S)} \cdot I^{\bar{S}(N^*)}$$

$1,665 = 1,643 \bullet 1,013$

Modificările absolute ale salariului mediu pe total, datorate influențelor factorilor sunt:

$$\Delta\bar{S}(S) = \frac{\Sigma S_1 N_1}{\Sigma N_1} - \frac{\Sigma S_0 N_1}{\Sigma N_1} = \Sigma S_1 N_1^* - \Sigma S_0 N_1^* = 3,83 - 2,33 = 1,5 \text{ mil. lei}$$

$$\Delta\bar{S}(N^*) = \frac{\Sigma S_0 N_1}{\Sigma N_1} - \frac{\Sigma S_0 N_0}{\Sigma N_0} = \Sigma S_0 N_1^* - \Sigma S_0 N_0^* = 2,33 - 2,3 = 0,03 \text{ mil. lei}$$

$$\Delta\bar{S}(S, N^*) = \Delta\bar{S}(S) + \Delta\bar{S}(N^*) \text{ mil. lei}$$

$1,53 = 1,5 + 0,03.$

Scanned with OKEN Scanner

Salariul mediu pe total S.C. ar fi crescut de 1,643 dacă s-ar fi modificat numai salariile la nivelul filialelor (această creştere ar fi însemnat, în expresie absolută, 1,5 mil. lei)

Dacă s-ar fi modificat numai structura salariaților, salariul mediu pe total S.C. ar fi crescut de 1,013 ori (adică cu 1,3% sau cu 30 mii lei).

7.5. Sisteme de ponderare folosite în construirea indicilor sintetici

De-a lungul timpului s-au conturat mai multe sisteme de ponderare. Alegerea ponderilor potrivite se face diferit, în funcție de conținutul indicelui în construirea căruia sunt implicate, de natura datelor existente în evidența curentă etc.

Sistemele de ponderare identificate şi aplicate până în prezent au încercat o caracterizare cât mai fidelă a dinamicii prețurilor, volumului fizic, circulației mărfurilor etc.

Dintre sistemele de ponderare folosite în mod curent în practica statistică amintim:

a) Sistemul de ponderare propus de statisticianul german ***Etienne Laspeyres*** în 1864, care foloseşte *ponderi din perioada de bază.*

În aceste condiții, indicii factoriali se vor scrie:

– indicele variabilei cantitative:

$$I_{1/0}^{f} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \qquad (7.77.)$$

– indicele variabilei calitative:

$$I_{1/0}^{x} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \qquad (7.78.)$$

b) Sistemul de pondere propus de statisticianul german ***Hermann Paasche*** în 1874 are la bază utilizarea *ponderilor din perioada curentă.*

Indicii factoriali construiți conform acestui sistem de ponderare vor fi:

– pentru variabila cantitativă:

$$I_{1/0}^{f} = \frac{\Sigma x_1 f_1}{\Sigma x_1 f_0} \qquad (7.79.)$$

– pentru variabila calitativă:

$$I_{1/0}^{x} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \qquad (7.80.)$$

Una din cerințele importante este îndeplinirea *testului de reversibilitate a factorilor,* adică produsul indicilor factoriali să fie egal cu indicele variației caracteristicii complexe. Pentru îndeplinirea acestei condiții, va trebui aleasă una

Scanned with OKEN Scanner

din variantele: indicele caracteristicii cantitative să fie construit ca indice Laspeyres iar indicele caracteristicii calitative – în sistem Paasche sau invers. Există, însă şi unele ţări care utilizează în construirea ambilor indici factoriali acelaşi sistem de ponderare (ori pe cel Laspeyres, ori pe cel Paasche). Fiecare variantă aleasă prezintă deopotrivă avantaje şi dezavantaje.

Au fost însă elaborate şi alte sisteme de ponderare, un fel de hibrid între sistemele Laspeyres şi Paasche, din care cele mai cunoscute sunt:

3) Indicele propus de ***Edgeworth***

Foloseşte ca pondere pentru variabila calitativă (preţul) fie suma ponderilor din cele două perioade, fie media aritmetică simplă a acestor ponderi. Are următoarea formă:

$$I_{1/0}^{x} = \frac{\Sigma x_1 (f_1 + f_0)}{\Sigma x_0 (f_1 + f_0)} = \frac{\Sigma x_1 \left(\frac{f_1 + f_0}{2} \right)}{\Sigma x_0 \left(\frac{f_1 + f_0}{2} \right)} \qquad (7.81.)$$

Acest sistem de ponderare prezintă dezavantajul că el se poate aplica numai la construirea indicelui de grup al factorului calitativ. Dacă s-ar construi şi indicele factorului cantitativ, folosind acest sistem de ponderare, el ar trebui să arate astfel:

$$I_{1/0}^{f} = \frac{\Sigma f_1 (x_1 + x_0)}{\Sigma f_0 (x_1 + x_0)} \qquad (7.82.)$$

care însă este lipsit de sens, deoarece valorile variabilei calitative nu pot fi însumate.

4) Sistemul de ponderare bazat pe utilizarea ponderilor din perioada de bază şi curentă, care stă la baza construirii indicelui „ideal" propus de ***I. Fischer***. Acest indice echivalează de fapt cu media geometrică a indicilor factorului în cauză în sistem Laspeyres şi Paasche:

– pentru factorul cantitativ:

$$I_{1/0}^{f} = \sqrt{I_{1/0(L)}^{f} \cdot I_{1/0(P)}^{f}} = \sqrt{\frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \cdot \frac{\Sigma x_1 f_1}{\Sigma x_1 f_0}} \qquad (7.83.)$$

– pentru factorul calitativ:

$$I_{1/0}^{x} = \sqrt{I_{1/0(L)}^{x} \cdot I_{1/0(P)}^{x}} = \sqrt{\frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \cdot \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1}} \qquad (7.84.)$$

Valoarea indicelui Fischer se va situa între valorile celor doi indici (Laspeyres şi Paasche) din care se calculează.

Scanned with OKEN Scanner

Sistemul de ponderare propus de Fischer are avantajul că indici factoriali calculaţi pe baza lui îndeplinesc condiţia de reversibilitate a factorilor. Din punct de vedere practic, însă, el prezintă şi un dezavantaj: acela că presupune existenţa datelor în evidenţa curentă privind toate variabilele implicate în construirea celor doi indici, şi din ambele perioade.

Acest sistem se foloseşte în construirea indicilor teritoriali, în comparaţiile pe plan naţional şi internaţional.

Pentru a se arăta gradul şi sensul influenţei ponderilor utilizate asupra indicelui de grup, se apelează la *relaţia lui Bortkiewicz*:

Pentru o variabilă calitativă, această relaţie este:

$$I^x_{i/0P} : I^x_{i/0L} = 1 + r_{ix_if} \cdot V_{ix} \cdot V_{if} \qquad (7.85.)$$

Notăm cu $P = r_{ix_if} \cdot V_{ix} \cdot V_{if}$. Pot exista următoarele cazuri:

• $P = 0$ când $r_{ix_if} = 0$ sau $V_{ix} = 0$ sau $V_{if} = 0$

– dacă $r_{ix_if} = 0$ însemnă că cele două variabile luate în calcul (i^x şi i^f) sunt independente;

– dacă $V_{ix} = 0$ înseamnă că toţi indicii individuali ai lui x sunt egali între ei, şi egali cu media;

– dacă $V_{if} = 0$ înseamnă că toţi indicii individuali ai lui f sunt egali între ei, şi egali cu media.

• $P > 0 \Rightarrow V_{ix_if} > 0$ când între cele două variabile există o legătură directă

• $P < 0 \Rightarrow V_{ix_if} < 0$ când între indicii individuali ai lui x şi f există o legătură inversă.

Vom spune că:

– dacă $I^x_{i/0P} = I^x_{i/0L}$ atunci sistemul de ponderare utilizat nu influenţează nivelul dinamicii fenomenului studiat

În acest caz $r_{ix_if} = 0$ sau $V_{ix} = 0$ sau $V_{if} = 0$

– dacă $I^x_{i/0P} > I^x_{i/0L}$ atunci cei trei factori sunt diferiţi de zero, $r_{ix_if} > 0 \Rightarrow p > 0$

– dacă $I^x_{i/0P} < I^x_{i/0L}$ atunci cei trei factori sunt diferiţi de zero, $r_{ix_if} < 0 \Rightarrow p < 0$

Scanned with OKEN Scanner

7.6. Determinarea contribuţiei factorilor la variaţia unui fenomen complex prin metoda indicilor

Am văzut în paragrafele precedente că majoritatea fenomenelor întâlnite în practica economică sunt fenomene complexe, a căror variaţie se află sub incidenţa a numeroşi factori. Ne este util să aflăm cu cât şi în ce sens s-a modificat valoarea producţiei, dar ar trebui să ştim şi sub influenţa căror factori s-a produs această modificare, precum şi în ce măsură au determinat aceştia modificarea, deplasarea produsă. Totodată, acelaşi fenomen se poate afla sub incidenţa mai multor tipuri de factori (de exemplu: producţia poate fi influenţată de preţuri şi cantităţi produse, sau de productivitatea muncii şi numărul de salariaţi etc.).

La rândul lor, unii factori pot fi şi ei influenţaţi de alţii (de pildă, productivitatea muncii (W) depinde de eficienţa fondurilor fixe (E_{FF}) şi de gradul de înzestrare a muncii cu fonduri fixe (Z) după relaţiile:

$$W = E_{FF} \cdot Z$$

unde $E_{FF} = \frac{Q}{FF}$ iar $Z = \frac{FF}{N}$, unde N = numărul de salariaţi

$$\Rightarrow W = \frac{Q}{FF} \cdot \frac{FF}{N} = \frac{Q}{N}$$

Aşadar, nivelul producţiei se va descompune în produsul a trei factori:

$$Q = W \cdot N = E_{FF} \cdot Z \cdot N$$

Pentru descompunerea variaţiei fenomenului complex pe factori de influenţă se foloseşte metoda indicilor.

Această metodă presupune descompunerea indicelui sintetic al variabilei complexe în produsul indicilor factorilor de influenţă (numită şi descompunere geometrică), precum şi descompunerea modificării absolute totale a fenomenului complex în suma modificărilor absolute datorate factorilor (numită şi descompunerea aritmetică sau analitică).

Cel mai des folosite metode de descompunere a variaţiei fenomenelor pe factori sunt:

a) metoda substituţiei în lanţ;

b) metoda restului nedescompus, cunoscută şi sub numele de metoda influenţei izolate a factorilor.

7.6.1. Metoda substituţiei în lanţ

Metoda presupune izolarea, pe rând, a fiecărui factor şi păstrarea variaţiei factorului a cărui influenţă ne interesează. În urma substituirii pe rând, a factorilor, se obţin indici factoriali, al căror număr este egal cu numărul factorilor. Ordinea în

Scanned with OKEN Scanner

care vor fi substituiți factorii este convențională. Se poate pleca de la ipoteza că dezvoltarea, modificarea fenomenelor se produce, în primul rând, datorită modificărilor cantitative, (acțiunii factorilor cantitativi) și apoi datorită factorilor calitativi (acțiunii factorilor calitativi). În construirea indicilor factoriali, atunci când are loc substituirea fiecărui factor, se folosesc sisteme de ponderare diferite.

Metoda presupune următoarele reguli:

– la substituirea primului factor (de regulă, cel cantitativ) se construiește indicele factorial al acestuia, folosind drept ponderi valorile celorlalți factori, nesubstituiți încă, la nivelul perioadei de bază;

– factorul substituit anterior se transformă în pondere la nivelul perioadei curente, intrând în construcția tuturor indicilor factoriali ce vor fi construiți.

Vom lua în considerare, în continuare, două variante, funcție de factorul cu care se începe substituția.

Fie y fenomenul complex și doi factori de influență ai acestuia x și f.

Varianta I (Figura 7.1.)

Se substituie întâi factorul cantitativ:

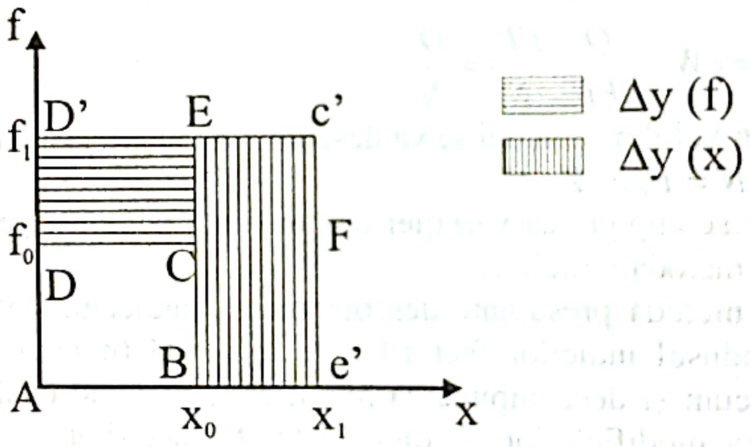


Figura 7.1.

Suprafața $ABCD$ echivalează cu nivelul caracteristicii complexe în perioada de bază

$$y_0 = x_0 f_0 \tag{7.86.}$$

iar suprafața $A'B'C'D'$ este egală cu nivelul caracteristicii complexe în perioada curentă

$$y_1 = x_1 f_1 \tag{7.87.}$$

Condițiile care se impun, așa cum reiese și din figură sunt:

$$y_1 > y_0, x_1 > x_0, f_1 > f_0 \tag{7.88.}$$

adică ambele variabile de influență și cea complexă au înregistrat creșteri.

Suprafața $BB'C'D'DC$ este egală cu diferența între suprafețele $A'B'C'D'$ și $ABCD$ și echivalează cu modificarea absolută a variabilei complexe y:

Scanned with OKEN Scanner

$$\Delta\Sigma y = \Sigma x_1 f_1 - \Sigma x_0 f_0 \qquad (7.89.)$$

Pentru fiecare factor, indicele şi modificarea absolută se calculează conform relaţiilor:

Factorul cantitativ

Descompunerea geometrică:

$$I_{1/0}^{\Sigma y(f)} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \qquad (7.90.)$$

Descompunerea analitică:

$$\Delta\Sigma y(f) = \Sigma x_0 f_1 - \Sigma x_0 f_0 \qquad (7.91.)$$

Factorul calitativ

Descompunerea geometrică:

$$I_{1/0}^{\Sigma y(x)} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \qquad (7.92.)$$

Descompunerea analitică:

$$\Delta\Sigma y(x) = \Sigma x_1 f_1 - \Sigma x_0 f_1 \qquad (7.93.)$$

$$I_{1/0}^{\Sigma y} = I^{\Sigma y(f)} \cdot I^{\Sigma y(x)} \qquad (7.94.)$$

$$\Delta\Sigma y = \Delta\Sigma y(f) + \Delta\Sigma y(x) \qquad (7.95.)$$

Suprafaţa *EC'FC* reprezintă restul nedescompus, adică partea din modificarea absolută totală a lui *y* care este determinată de influenţa concomitentă a ambilor factori (*x* şi *f*). În această variantă, restul nedescompus este atribuit în întregime factorului *x*.

Varianta II (Figura 7.2.)

Se substituie întâi factorul calitativ:

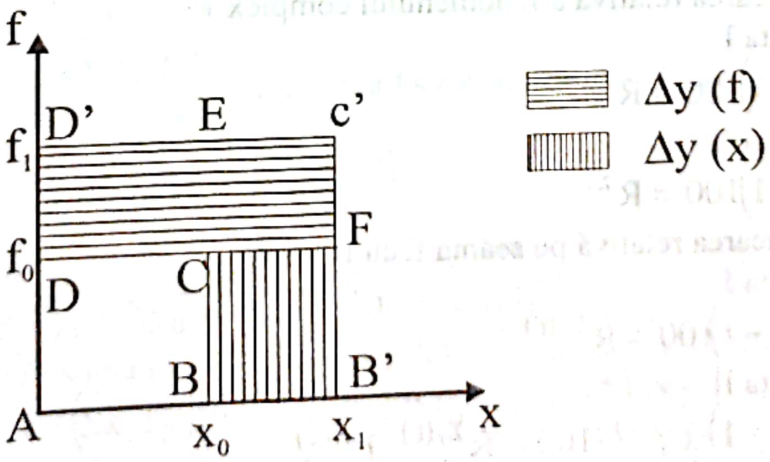


Figura 7.2.

Scanned with OKEN Scanner

Indicii şi modificările absolute se calculează conform relaţiilor:

Factorul calitativ

Descompunerea geometrică:

$$I^{\Sigma y(x)} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \qquad (7.96.)$$

Descompunerea analitică:

$$\Delta\Sigma y(x) = \Sigma x_1 f_0 - \Sigma x_0 f_0 \qquad (7.97.)$$

Factorul cantitativ

Descompunerea geometrică:

$$I^{\Sigma y(f)} = \frac{\Sigma x_1 f_1}{\Sigma x_1 f_0} \qquad (7.98.)$$

Descompunerea analitică:

$$\Delta\Sigma y(f) = \Sigma x_1 f_1 - \Sigma x_1 f_0 \qquad (7.99.)$$

$$I_{1/0}^{\Sigma y} = I^{\Sigma y(x)} \cdot I^{\Sigma y(f)} \qquad (7.100.)$$

$$\Delta\Sigma y = \Delta\Sigma y(x) + \Delta\Sigma y(f) \qquad (7.101.)$$

Restul nedescompus este atribuit în întregime, în această variantă, factorului *f*.

Cu ajutorul acestei metode se poate determina şi influenţa factorilor asupra modificării relative:

Modificarea relativă a fenomenului complex *y*:

Varianta I

$$\left(I^{\Sigma y} - 1\right)100 = R^{\Sigma y} \qquad (7.102.)$$

Varianta II

$$\left(I^{\Sigma y} - 1\right)100 = R^{\Sigma y} \qquad (7.103.)$$

Modificarea relativă pe seama factorului *f*:

Varianta I

$$\left(I^{\Sigma y(f)} - 1\right)100 = R^{\Sigma y(f)} \qquad (7.104.)$$

Varianta II

$$\left(I^{\Sigma y(f)} - 1\right) \cdot I^{\Sigma y(x)} \cdot 100 = R^{\Sigma y(f)} \cdot I^{\Sigma y(x)} \qquad (7.105.)$$

Scanned with OKEN Scanner

Modificarea relativă pe seama factorului *x*:

Varianta I

$$\left(I^{\Sigma y(x)} - 1\right) \cdot I^{\Sigma y(f)} \cdot 100 = R^{\Sigma y(x)} \cdot I^{\Sigma y(f)} \qquad (7.106.)$$

Varianta II

$$\left(I^{\Sigma y(x)} - 1\right) 100 = R^{\Sigma y(x)} \qquad (7.107.)$$

Metoda substituției în lanț se poate aplica și atunci când fenomenul complex este influențat de doi factori. Vom lua, de exemplu, 3 factori (din care unul cantitativ, unul structural și unul calitativ) ce determină un fenomen complex *y*.

Presupunem că factorul cantitativ determină variația variabilei complexe *y* atât ca volum, cât și ca structură. Cel mai adesea se substituie întâi factorul cantitativ, apoi cel structural și ultimul – cel calitativ.

Fenomenul complex va fi scris sub forma:

$$\Sigma y = \left(\Sigma \times f^*\right)\Sigma f \qquad (7.108.)$$

Indicele general al variabilei complexe:

$$I_{1/0}^{\Sigma y} = \frac{\Sigma y_1}{\Sigma y_0} = \frac{\left(\Sigma x_1 f_1^*\right)\Sigma f_1}{\left(\Sigma x_0 f_0^*\right)\Sigma f_0} \qquad (7.109.)$$

se descompune în 3 indici factoriali:

$$I_{1/0}^{\Sigma y(f)} = \frac{\left(\Sigma x_0 f_0^*\right)\Sigma f_1}{\left(\Sigma x_0 f_0^*\right)\Sigma f_0} \text{ (influența factorului cantitativ)} \qquad (7.110.)$$

$$I_{1/0}^{\Sigma y(f^*)} = \frac{\left(\Sigma x_0 f_1^*\right)\Sigma f_1}{\left(\Sigma x_0 f_0^*\right)\Sigma f_1} \text{ (influența factorului structural)} \qquad (7.111.)$$

$$I_{1/0}^{\Sigma y(x)} = \frac{\left(\Sigma x_1 f_1^*\right)\Sigma f_1}{\left(\Sigma x_0 f_1^*\right)\Sigma f_1} \text{ (influența factorului calitativ)} \qquad (7.112.)$$

$$I_{1/0}^{\Sigma y(x,f,f^*)} = I_{1/0}^{\Sigma y(x)} \cdot I_{1/0}^{\Sigma y(f)} \cdot I_{1/0}^{\Sigma y(f^*)} \qquad (7.113.)$$

Modificarea absolută a variabilei complexe:

$$\Delta\Sigma y = \Sigma y_1 - \Sigma y_0 = \left(\Sigma x_1 f_1^*\right)\Sigma f_1 - \left(\Sigma x_0 f_0^*\right)\Sigma f_0 \qquad (7.114.)$$

se descompune în suma modificărilor absolute datorate celor 3 factori:

$$\Delta\Sigma y(f) = \left(\Sigma x_0 f_0^*\right)\Sigma f_1 - \left(\Sigma x_0 f_0^*\right)\Sigma f_0 = \left(\Sigma x_0 f_0^*\right)\Delta\Sigma f \qquad (7.115.)$$

$$\Delta\Sigma y(f^*) = \left(\Sigma x_0 f_1^*\right)\Sigma f_1 - \left(\Sigma x_0 f_0^*\right)\Sigma f_1 = \left(\Sigma x_0 \Delta f^*\right)\Sigma f_1 \qquad (7.116.)$$

$$\Delta\Sigma y(x) = \left(\Sigma x_1 f_1^*\right)\Sigma f_1 - \left(\Sigma x_0 f_1^*\right)\Sigma f_1 = \left(\Sigma \Delta x f_1^*\right)\Sigma f_1 \qquad (7.117.)$$

$$\Delta\Sigma y(x,f,f^*) = \Delta\Sigma y(x) + \Delta\Sigma y(f) + \Delta\Sigma y(f^*) \qquad (7.118.)$$

Scanned with OKEN Scanner

Metoda substituției în lanț se aplică, în practică, datorită simplicității ei, mai ales atunci când dorim să studiem variația fenomenelor pe perioade scurte de timp.

Exemplul 7.9. Despre productivitatea muncii și numărul salariaților din două secții ale unei întreprinderi, se cunosc următoarele date, în ianuarie și mai 2006:

Tabelul 7.8.

Secția	Productivitatea muncii (mil.lei/pers.)		Nr. salariați (pers.)	
	ian. 2006	mai 2006	ian. 2006	mai 2006
0	1	2	3	4
A	3,00	6,5	20	25
B	4,25	5,0	40	30
Total	–	–	60	55

Se fac următoarele notații:

w_0, w_1 = productivitatea muncii la nivel de secție, în perioada de bază și curentă;

N_0, N_1 = nr. de salariați ai fiecărei secții în perioada de bază și curentă;

q_0, q_1 = valoarea producției realizată de fiecare secție în perioada de bază și curentă.

$$W_0 = \frac{q_0}{N_0} \Rightarrow q_0 = W_0 N_0 \rightarrow q_0^A = W_0^A N_0^A = 3 \cdot 20 = 60 \text{ mil. lei}$$

$$\rightarrow q_0^B = W_0^B \cdot N_0^B = 4{,}25 \cdot 40 = 170 \text{ mil. lei}$$

$$\Rightarrow \Sigma q_0 = \Sigma W_0 N_0 = 60 + 170 = 230 \text{ mil lei}$$

$$W_1 = \frac{q_1}{N_1} \Rightarrow q_1 = W_1 N_1 \rightarrow q_1^A = W_1^A N_1^A = 6{,}5 \cdot 25 = 162{,}5 \text{ mil. lei}$$

$$\rightarrow q_1^B = W_1^B \cdot N_1^B = 5 \cdot 30 = 150 \text{ mil. lei}$$

$$\Rightarrow \Sigma q_1 = \Sigma W_1 N_1 = 162{,}5 + 150 = 312{,}5 \text{ mil. lei}$$

Pentru separarea dinamicii și modificării absolute a producției pe cei doi factori de influență (W, N) se aplică metoda substituției în lanț. Considerăm că primul se substituie factorul cantitativ (N):

$$I^{\Sigma q} = \frac{\Sigma q_1}{\Sigma q_0} = \frac{\Sigma W_1 N_1}{\Sigma W_0 N_0} = \frac{312{,}5}{230} = 1{,}36\,(136\%)$$

$$\Delta \Sigma q = \Sigma q_1 - \Sigma q_0 = \Sigma W_1 N_1 - \Sigma W_0 N_0 = 312{,}5 - 230 = 82{,}5 \text{ mil. lei}$$

$$I^{\Sigma q(N)} = \frac{\Sigma W_0 N_1}{\Sigma W_0 N_0} = \frac{3 \cdot 25 + 4{,}25 \cdot 30}{230} = \frac{75 + 127{,}5}{230} = \frac{202{,}5}{230} = 0{,}88$$

Scanned with OKEN Scanner

$$\Delta\Sigma q(N) = \Sigma W_0 N_1 - \Sigma W_0 N_0 = 202{,}5 - 230 = -27{,}5 \text{ mil. lei}$$

$$I^{\Sigma q(W)} = \frac{\Sigma W_1 N_1}{\Sigma W_0 N_1} = \frac{312{,}5}{202{,}5} = 1{,}54\,(154\%)$$

$$\Delta\Sigma q(W) = \Sigma W_1 N_1 - \Sigma W_0 N_1 = 312{,}5 - 202{,}5 = 110 \text{ mil. lei}$$

$$I^{\Sigma q(W,N)} = I^{\Sigma q(W)} \cdot I^{\Sigma q(N)}$$

$$\Delta\Sigma q(W,N) = \Delta\Sigma q(W) + \Delta\Sigma q(N)$$

Aşadar, valoarea producţiei întreprinderii a crescut în mai faţă de ianuarie 2006 de 1,36 ori (adică cu 36%, sau cu 82,5 mil. lei). Dacă s-ar fi modificat doar numărul salariaţilor, iar productivitatea muncii ar fi rămas constantă, producţia totală ar fi scăzut cu 12% (sau cu 27,5% mil. lei), în timp ce, sub influenţa modificării productivităţii muncii (numărul salariaţilor – rămânând neschimbat), valoarea producţiei ar fi crescut în perioada analizată de 1,54 ori (sau cu 54%, adică cu 110 mil. lei).

7.6.2. Metoda influenţelor izolate ale factorilor (metoda restului nedescompus)

Prin această metodă se acordă aceeaşi importanţă tuturor factorilor. Spre deosebire de metoda substituţiei în lanţ, indicii factoriali se construiesc pe baza aceluiaşi sistem de ponderare (Laspeyres). Ponderile vor fi, deci, cele din perioada de bază, variaţia fiecărui factor are loc în condiţiile în care celălalt sau ceilalţi factori rămân nemodificaţi (la nivelul perioadei de bază).

Spunem că indicii astfel calculaţi vor reflecta influenţa izolată (sau independentă) a factorilor.

Fie fenomenul complex y, determinat de cei doi factori: f (factor cantitativ) şi x (factor calitativ):

$$y = xf \tag{7.119.}$$

Indicele influenţei izolate a lui f se calculează conform relaţiei:

$$I^{\Sigma y(f/x_0)} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \tag{7.120.}$$

Indicele influenţei izolate a lui x este dat de relaţia:

$$I^{\Sigma y(x/f_0)} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \tag{7.121.}$$

Modificările absolute corespunzătoare vor fi:

$$\Delta\Sigma y(f/x_0) = \Sigma x_0 f_1 - \Sigma x_0 f_0 \tag{7.122.}$$

$$\Delta\Sigma y(x/f_0) = \Sigma x_1 f_0 - \Sigma x_0 f_0 \tag{7.123.}$$

Scanned with OKEN Scanner

Făcând produsul indicilor factoriali calculaţi, acesta nu este egal cu indicele variaţiei totale a fenomenului complex *y*:

$$I^{\Sigma y(f/x_0)} \cdot I^{\Sigma y(x/f_0)} \neq I^{\Sigma y} \qquad (7.124.)$$

$$\frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \cdot \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \neq \frac{\Sigma x_1 f_1}{\Sigma x_0 f_0} \qquad (7.125.)$$

Testul de reversibilitate a factorilor nu mai este îndeplinit, deoarece în construirea indicilor factoriali se foloseşte acelaşi sistem de ponderare şi deci nu este reflectată şi influenţa variaţiei ponderilor.

La fel, suma modificărilor absolute ale factorilor nu mai este egală cu modificarea absolută totală a variabilei complexe:

$$\Delta\Sigma y(f/x_0) + \Delta\Sigma y(x/f_0) \neq \Delta\Sigma y \qquad (7.126.)$$

$$(\Sigma x_0 f_1 - \Sigma x_0 f_0) + (\Sigma x_1 f_0 - \Sigma x_0 f_0) \neq \Sigma x_1 f_1 - \Sigma x_0 f_0 \qquad (7.127.)$$

Mărimea cu care diferă produsul indicilor factoriali de indicele general este denumită ***rest nedescompus*** şi reprezintă variaţia fenomenului complex datorată acţiunii (variaţiei) concomitente a celor doi factori. Geometric, mărimea restului nedescompus este egală cu aria suprafeţei CEC'D din Figura 7.3.

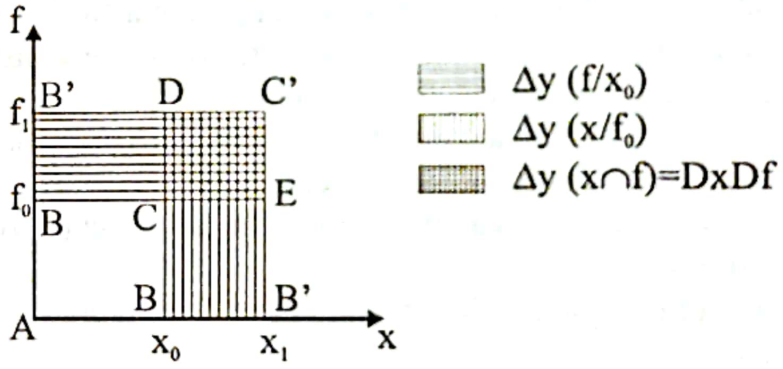


Figura 7.3.

Aria suprafeţei CEC'D este egală cu produsul modificărilor $(x_1 - x_0) \cdot (y_1 - y_0)$. Pe lângă indicii influenţelor independente ale factorilor, se construieşte şi un indice ce reflectă influenţa comună, simultană a ambilor factori. Acesta se calculează ca raport între indicele factorului calitativ de tip Paasche şi indicele factorului calitativ de tip Laspeyres. În felul acesta indicele astfel determinat reflectă şi influenţa variaţiei ponderilor folosite:

$$I^{\Sigma y(x \cap f)} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} : \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \qquad (7.128.)$$

Analog, se determină şi modificarea absolută a variabilei complexe *y*, datorată variaţiei simultane a celor doi factori:

Scanned with OKEN Scanner

$$\Delta\Sigma y(x \cap f) = (\Sigma x_1 f_1 - \Sigma x_0 f_1) - (\Sigma x_1 f_0 - \Sigma x_0 f_0) \qquad (7.129.)$$

Aşadar, indicele variaţiei caracteristicii complexe *y* se descompune în produsul a 3 indicatori factoriali (descompunerea geometrică):

$$I^{\Sigma y(x,f)} = I^{\Sigma y(f/x_0)} \cdot I^{\Sigma y(x/f_0)} \cdot I^{\Sigma y(x \cap f)} \qquad (7.130)$$

$$\frac{\Sigma x_1 f_1}{\Sigma x_0 f_0} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0} \cdot \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} \cdot \left(\frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \cdot \frac{\Sigma x_0 f_0}{\Sigma x_1 f_0} \right) \qquad (7.131)$$

Modificarea absolută totală a variabilei *y* se descompune în suma a 3 modificări absolute (descompunerea analitică):

$$\Delta\Sigma y(x,f) = \Delta\Sigma y(f/x_0) + \Delta\Sigma y(x/f_0) + \Delta\Sigma y(x \cap f) \qquad (7.132.)$$

$$\Sigma x_1 f_1 - \Sigma x_0 f_0 = (\Sigma x_0 f_1 - \Sigma x_0 f_0) + (\Sigma x_1 f_0 - \Sigma x_0 f_0) + [(\Sigma x_1 f_1 - \Sigma x_0 f_1) - (\Sigma x_1 f_0 - \Sigma x_0 f_0)] \qquad (7.133.)$$

sau

$$\Delta\Sigma y(x,f) = \Sigma x_0 \Delta f + \Sigma f_0 \Delta x + \Sigma \Delta x \Delta f \qquad (7.134.)$$

Pentru a calcula influenţa *totală* a fiecărui factor, este necesar ca şi restul nedescompus să fie împărţit pe cei doi factori de influenţă:

Separarea restului nedescompus pe factori se poate realiza în 3 moduri:

– restul nedescompus se atribuie, în întregime, unui singur factor (lui *x* sau lui *f*). În acest caz, metoda coincide cu metoda substituţiei în lanţ. Atribuirea se va face factorului calitativ (în varianta I a metodei substituţiei în lanţ) sau factorului cantitativ (în varianta II a metodei substituţiei în lanţ);

– restul nedescompus se atribuie în mod egal celor doi factori.

Indicele influenţei totale a factorului *f* va fi:

$$I^{\Sigma y(f)} = I^{\Sigma y(f/x_0)} \cdot \sqrt{I^{\Sigma y(x \cap f)}} \qquad (7.135.)$$

Indicele influenţei totale a factorului *x* va fi:

$$I^{\Sigma y(x)} = I^{\Sigma y(x/f_0)} \cdot \sqrt{I^{\Sigma y(x \cap f)}} \qquad (7.136.)$$

Modificările absolute ce reflectă influenţele totale ale factorilor vor fi:

$$\Delta\Sigma y(f) = \Delta\Sigma y(f/x_0) + \frac{\Delta\Sigma y(x \cap f)}{2} \qquad (7.137.)$$

$$\Delta\Sigma y(x) = \Delta\Sigma y(x/f_0) + \frac{\Delta\Sigma y(x \cap f)}{2} \qquad (7.138.)$$

– restul nedescompus se atribuie ambilor factori, proporţional cu influenţa izolată independentă, a fiecăruia.

$$I^{\Sigma y(f)} = I^{\Sigma y(f/x_0)} \cdot \sqrt{\frac{I^{\Sigma y(f/x_0)}}{I^{\Sigma y(x/f_0)}} \cdot I^{\Sigma y(x \cap f)}} \qquad (7.139.)$$

Scanned with OKEN Scanner

$$I^{\Sigma y(x)} = I^{\Sigma y(x/f_0)} \cdot \sqrt{\frac{I^{\Sigma y(x/f_0)}}{I^{\Sigma y(f/x_0)}} \cdot I^{\Sigma y(x \cap f)}} \quad (7.140.)$$

$$I^{\Sigma y(x)} \cdot I^{\Sigma y(f)} = I^{\Sigma y} \quad (7.141.)$$

$$\Delta\Sigma y(f) = \Delta\Sigma y(f/x_0) + \frac{\Delta\Sigma y(f/x_0)}{\Delta\Sigma y(f/x_0) + \Delta\Sigma y(x/f_0)} \cdot \Delta\Sigma y(x \cap f) \quad (7.142.)$$

$$\Delta\Sigma y(x) = \Delta\Sigma y(x/f_0) + \frac{\Delta\Sigma y(x/f_0)}{\Delta\Sigma y(x/f_0) + \Delta\Sigma y(f/x_0)} \cdot \Delta\Sigma y(x \cap f) \quad (7.143.)$$

$$\Delta\Sigma y(f) + \Delta\Sigma y(x) = \Delta\Sigma y \quad (7.144.)$$

Pe baza modificărilor absolute ale factorilor se determină ponderea contribuției fiecărui factor la modificarea absolută totală a variabilei complexe:

$$\frac{\Delta\Sigma y(f)}{\Delta\Sigma y(x,f)} \cdot 100; \qquad \frac{\Delta\Sigma y(x)}{\Delta\Sigma y(x,f)} \cdot 100 \quad (7.145.)$$

Metoda influențelor izolate permite, spre deosebire de metoda substituției în lanț, identificarea mai exactă a contribuției factorilor de influență la variația totală a unui fenomen complex, însă ea devine mai greoaie, mai dificil de aplicat atunci când numărul factorilor de influență este mai mare.

Exemplul 7.10. Despre două secții ale unei întreprinderi cu profil industrial se cunosc următoarele date, privind lunile ianuarie și februarie 2006:

Tabelul 7.9

Secția	Eficiența fondurilor fixe		Valoarea fondurilor fixe	
	ian. 2006	feb. 2006	ian. 2006	feb. 2006
0	1	2	3	4
A	13	30	10	20
B	20	25	15	20

Se fac următoarele notații:

Ef_0, Ef_1 = eficiența fondurilor fixe la nivelul secțiilor în perioada de bază (ianuarie) și cea curentă (februarie);

FF_0, FF_1 = valoarea fondurilor fixe pentru fiecare secție în perioada de bază și cea curentă;

q_0, q_1 = valoarea producției pe secții, în perioada de bază și cea curentă.

$$Ef_0 = \frac{q_0}{FF_0} \Rightarrow q_0 = Ef_0 \cdot FF_0 \rightarrow q_0^A = Ef_0^A \cdot FF_0^A = 13 \cdot 10 = 130 \text{ mld. lei}$$

$$\rightarrow q_0^B = Ef_0^B \cdot FF_0^B = 20 \cdot 15 = 300 \text{ mld. lei}$$

Scanned with OKEN Scanner

$$Ef_1 = \frac{q_1}{FF_1} \Rightarrow q_1 = Ef_1 \cdot FF_1 \rightarrow q_1^A = Ef_1^A \cdot FF_1^A = 30 \cdot 20 = 600 \text{ mld. lei}$$

$$\rightarrow q_1^B = Ef_1^B \cdot FF_1^B = 25 \cdot 20 = 500 \text{ mld. lei}$$

Pe total, producțiile vor fi:

$$\Sigma q_0 = 130 + 300 = 430 \text{ mld. lei}$$

$$\Sigma q_1 = 600 + 500 = 1100 \text{ mld. lei}$$

$$I^{\Sigma q} = \frac{\Sigma q_1}{\Sigma q_0} = \frac{\Sigma Ef_1 \cdot FF_1}{\Sigma Ef_0 \cdot FF_0} = \frac{1100}{430} = 2{,}558$$

$$\Delta\Sigma q = \Sigma q_1 - \Sigma q_0 = \Sigma Ef_1 FF_1 - \Sigma Ef_0 FF_0 = 1100 - 430 = 670 \text{ mld. lei}$$

Influențele izolate ale eficienței fondurilor fixe şi valorii fondurilor fixe asupra variației producției sunt:

$$I^{\Sigma q(Ef/FF_0)} = \frac{\Sigma Ef_1 FF_0}{\Sigma Ef_0 FF_0} = \frac{30 \cdot 10 + 25 \cdot 15}{430} = \frac{675}{430} = 1{,}57$$

$$\Delta\Sigma q(Ef/FF_0) = \Sigma Ef_1 FF_0 - \Sigma Ef_0 FF_0 = 675 - 430 = 245 \text{ mld. lei}$$

$$I^{\Sigma q(FF/Ef_0)} = \frac{\Sigma Ef_0 FF_1}{\Sigma Ef_0 FF_0} = \frac{13 \cdot 20 + 20 \cdot 20}{430} = \frac{660}{430} = 1{,}535$$

$$\Delta\Sigma q(FF/Ef_0) = \Sigma Ef_0 FF_1 - \Sigma Ef_0 FF_0 = 660 - 430 = 230 \text{ mld. lei}$$

Rezultatele obținute arată că:

În condițiile menținerii fondurilor fixe la valoarea din perioada de bază, variația eficienței fondurilor fixe ar fi determinat o creştere a valorii producției de 1,57 ori (sau cu 57%, adică cu 245 mld.lei).

Producția ar fi crescut în perioada curentă, fată de perioada de bază de 1,535 ori (cu 53,5%, sau cu 230 mld. lei), sub influența modificării valorii fondurilor fixe, în condițiile în care eficiența fondurilor fixe ar fi rămas constantă.

Restul nedescompus, datorat influenței ambilor factori, va fi:

$$I^{\Sigma q(Ef \cap FF)} = \frac{\Sigma Ef_1 FF_1}{\Sigma Ef_0 FF_1} : \frac{\Sigma Ef_1 FF_0}{\Sigma Ef_0 FF_0} = \frac{1100}{660} \cdot \frac{430}{675} = 1{,}06$$

$$\Delta\Sigma q(Ef \cap FF) = (\Sigma Ef_1 FF_1 - \Sigma Ef_0 FF_1) - (\Sigma Ef_1 FF_0 - \Sigma Ef_0 FF_0) =$$

$$= (1100 - 660) - (675 - 430) = 440 - 245 = 195 \text{ mld. lei}$$

Sub influența modificării atât a eficienței fondurilor fixe, cât şi a valorii fondurilor fixe, producția a crescut în perioada curentă față de cea de bază de 1,06 ori (adică cu 6% sau cu 195 mld. lei).

Scanned with OKEN Scanner

Descompunerea geometrică

$$I^{\Sigma q(Ef/FF_0)} \cdot I^{\Sigma q(FF/Ef_0)} \cdot I^{\Sigma q(FF \cap Ef)} = I^{\Sigma q(Ef,FF)}$$

$$1{,}57 \cdot 1{,}535 \cdot 1{,}06 \approx 2{,}55$$

Descompunerea analitică:

$$\Delta\Sigma q(Ef/FF_0) + \Delta\Sigma q(FF/Ef_0) + \Delta\Sigma q(FF \cap Ef) = \Delta\Sigma q(Ef,FF)$$

$$245 + 230 + 195 = 670 \text{ mld. lei}$$

Influența totală a factorilor va fi:

$$I^{\Sigma q(Ef)} = I^{\Sigma q(Ef/FF_0)} \cdot \sqrt{\frac{I^{\Sigma q(Ef/FF_0)}}{I^{\Sigma q(FF/Ef_0)}} \cdot I^{\Sigma q(Ef \cap FF)}} = 1{,}635$$

$$I^{\Sigma q(FF)} = I^{\Sigma q(FF/Ef_0)} \cdot \sqrt{\frac{I^{\Sigma q(FF/Ef_0)}}{I^{\Sigma q(Ef/FF_0)}} \cdot I^{\Sigma q(Ef \cap FF)}} = 1{,}563$$

$$I^{\Sigma q(Ef)} \cdot I^{\Sigma q(FF)} = 1{,}635 \cdot 1{,}563 = 2{,}55 = I^{\Sigma q}$$

$$\Delta\Sigma q(Ef) = \Delta\Sigma q(Ef/FF_0) + \frac{\Delta\Sigma q(Ef/FF_0)}{\Delta\Sigma q(Ef/FF_0) + \Delta\Sigma q(FF/Ef_0)} \cdot \Delta\Sigma q(Ef \cap FF) =$$

$$= 345{,}6 \text{ mld. lei}$$

$$\Delta\Sigma q(FF) = \Delta\Sigma q(FF/Ef_0) + \frac{\Delta\Sigma q(FF/Ef_0)}{\Delta\Sigma q(FF/Ef_0) + \Delta\Sigma q(Ef/FF_0)} \cdot \Delta\Sigma q(Ef \cap FF) =$$

$$= 324{,}4 \text{ mld. lei}$$

$$\Delta\Sigma q(Ef) + \Delta\Sigma q(FF) = 345{,}6 + 324{,}4 = 670 \text{ mld. lei}$$

Ponderea influenței factorilor la modificarea absolută totală a producției va fi:

$$\frac{\Delta\Sigma q(Ef)}{\Delta\Sigma q(Ef,FF)} \cdot 100 = \frac{345{,}6}{670} \cdot 100 = 51{,}6\%$$

$$\frac{\Delta\Sigma q(FF)}{\Delta\Sigma q(Ef,FF)} \cdot 100 = \frac{324{,}4}{670} \cdot 100 = 48{,}4\%$$

7.7. Serii de indici statistici

Am văzut că indicele statistic permite efectuarea unor comparații între două niveluri ale aceluiași fenomen, înregistrate în unități diferite de timp, spațiu sau față de plan.

Scanned with OKEN Scanner

Deseori, evoluţia unui fenomen în timp poate fi caracterizată printr-o suită de indici calculaţi pentru orizontul de timp luat în studiu (de exemplu, evoluţia indicelui PIB al României în perioada 1990-2005).

În asemenea cazuri, spunem că avem de-a face cu serii cronologice, alcătuite din indici.

La construirea acestor serii, trebuie stabilite două probleme mai importante:

a) baza de raportare;

b) ponderile folosite (acolo unde e cazul).

a) **baza de raportare**

Sub acest aspect, seriile de indici vor fi:

– cu bază fixă;

– cu bază în lanţ.

b) **ponderile folosite**

Ponderile se utilizează în cazul seriilor alcătuite din indici sintetici, care se referă la variabile cu valori neînsumabile direct. Ponderile ce pot fi utilizate sunt de două tipuri:

– ponderi constante – atunci când se folosesc aceleaşi ponderi în toată seria;

– ponderi variabile – atunci când ponderile se modifică odată cu modificarea bazei de raportare.

În funcţie de modul cum se alege baza de raportare şi sistemul de ponderare, există următoarele tipuri de serii de indici:

a) serii de indici de grup cu bază fixă şi ponderi constante;

b) serii de indici de grup cu bază în lanţ şi ponderi constante;

c) serii de indici de grup cu bază în lanţ şi ponderi variabile.

Notăm cu (x_i) valorile caracteristicii calitative şi cu (f_i) valorile caracteristicii cantitative.

a) **Serii de indici de grup cu bază fixă şi ponderi constante**

Se consideră ponderile la nivelul perioadei de bază

• Indicii construiţi pentru variabila x_i vor forma o serie de forma:

$$I^x_{i/o} = \left(\frac{\sum x_i f_0}{\sum x_0 f_0} \right)_{i=\overline{1,n}} \qquad (7.146.)$$

care se poate detalia astfel:

$$I^x_{1/0} = \frac{\sum x_1 f_0}{\sum x_o f_0}; I^x_{2/0} = \frac{\sum x_2 f_0}{\sum x_0 f_0}; I^x_{3/0} = \frac{\sum x_3 f_0}{\sum x_o f_0}; \ldots; I^x_{n/0} = \frac{\sum x_n f_0}{\sum x_o f_0} \qquad (7.147.)$$

Scanned with OKEN Scanner

• Indicii construiți pentru variabila f_i formează o serie de forma:

$$I^f_{1/0} = \left(\frac{\sum x_o f_i}{\sum x_0 f_0} \right)_{i=\overline{1,n}} \quad (7.148.)$$

care se detaliază astfel:

$$I^f_{1/0} = \frac{\sum x_0 f_1}{\sum x_o f_0}; I^f_{2/0} = \frac{\sum x_0 f_2}{\sum x_o f_0}; I^f_{3/0} = \frac{\sum x_0 f_3}{\sum x_o f_0}; ...; I^f_{n/0} = \frac{\sum x_0 f_n}{\sum x_o f_0} \quad (7.149.)$$

b) Serii de indici de grup cu bază în lanț și ponderi constante

• Pentru variabila x_i:

Ponderile se pot alege:

– la nivelul perioadei de bază, caz în care seria va avea forma:

$$I^x_{i/i-1} = \left(\frac{\sum x_i f_0}{x_{i-1} f_0} \right)_{i=\overline{1,n}} \quad (7.150.)$$

care, în forma desfășurată, se scrie:

$$I^x_{1/0} = \frac{\sum x_1 f_0}{\sum x_o f_0}; I^x_{2/1} = \frac{\sum x_2 f_0}{\sum x_1 f_0}; I^x_{3/2} = \frac{\sum x_3 f_0}{\sum x_2 f_0}; ...; I^x_{n/n-1} = \frac{\sum x_n f_0}{\sum x_{n-1} f_0} \quad (7.151.)$$

– la nivelul perioadei curente, caz în care seria va avea forma:

$$I^x_{i/i-1} = \left(\frac{\sum x_i f_n}{\sum x_{i-1} f_n} \right)_{i=\overline{1,n}} \quad (7.152.)$$

$$I^x_{1/0} = \frac{\sum x_1 f_n}{\sum x_o f_n}; I^x_{2/1} = \frac{\sum x_2 f_n}{\sum x_1 f_n}; I^x_{3/2} = \frac{\sum x_3 f_n}{\sum x_2 f_n}; ...; I^x_{n/n-1} = \frac{\sum x_n f_n}{\sum x_{n-1} f_n} \quad (7.153.)$$

Între termenii fiecăreia dintre aceste serii se verifică relația dintre indicii cu bază în lanț și cei cu bază fixă, adică:

$$\frac{\sum x_1 f_0}{\sum x_0 f_0} \cdot \frac{\sum x_2 f_0}{\sum x_1 f_0} \cdot \frac{\sum x_3 f_0}{\sum x_2 f_0} \cdot ... \cdot \frac{\sum x_n f_0}{\sum x_{n-1} f_0} = \frac{\sum x_n f_0}{\sum x_0 f_0} \quad (7.154.)$$

$$\frac{\sum x_1 f_n}{\sum x_0 f_n} \cdot \frac{\sum x_2 f_n}{\sum x_1 f_n} \cdot \frac{\sum x_3 f_n}{\sum x_2 f_n} \cdot ... \cdot \frac{\sum x_n f_n}{\sum x_{n-1} f_n} = \frac{\sum x_n f_n}{\sum x_0 f_n} \quad (7.155.)$$

• Pentru caracteristica f_i, avem:

– ponderi la nivelul perioadei de bază:

$$I^f_{i/i-1} = \left(\frac{\sum x_0 f_i}{\sum x_0 f_{i-1}} \right)_{i=\overline{1,n}} \quad (7.156.)$$

Scanned with OKEN Scanner

$$I^{f}_{1/0}=\frac{\sum x_0 f_1}{\sum x_0 f_0};I^{f}_{2/1}=\frac{\sum x_0 f_2}{\sum x_0 f_1};\ I^{f}_{3/2}=\frac{\sum x_0 f_3}{\sum x_0 f_2};\ ...;I^{f}_{n/n-1}=\frac{\sum x_0 f_n}{\sum x_0 f_{n-1}} \quad (7.157.)$$

– ponderi la nivelul perioadei curente:

$$I^{f}_{i/i-1}=\left(\frac{\sum x_n f_i}{\sum x_n f_{i-1}}\right)_{i=\overline{1,n}} \quad (7.158.)$$

$$I^{f}_{1/0}=\frac{\sum x_n f_1}{\sum x_n f_0};I^{f}_{2/1}=\frac{\sum x_n f_2}{\sum x_n f_1};\ I^{f}_{3/2}=\frac{\sum x_n f_3}{\sum x_n f_2};\ ...;I^{f}_{n/n-1}=\frac{\sum x_n f_n}{\sum x_n f_{n-1}} \quad (7.159.)$$

Şi în acest caz se verifică relaţia dintre indicii cu bază în lanţ şi ultimul cu bază fixă, şi anume:

$$\frac{\sum x_0 f_1}{\sum x_0 f_0}\cdot\frac{\sum x_0 f_2}{\sum x_0 f_1}\cdot\frac{\sum x_0 f_3}{\sum x_0 f_2}\cdot...\cdot\frac{\sum x_0 f_n}{\sum x_0 f_{n-1}}=\frac{\sum x_0 f_n}{\sum x_0 f_0} \quad (7.160.)$$

$$\frac{\sum x_n f_1}{\sum x_n f_0}\cdot\frac{\sum x_n f_2}{\sum x_n f_1}\cdot\frac{\sum x_n f_3}{\sum x_n f_2}\cdot...\cdot\frac{\sum x_n f_n}{\sum x_n f_{n-1}}=\frac{\sum x_n f_n}{\sum x_n f_0} \quad (7.161.)$$

Seriile de indici cu bază în lanţ şi ponderi constante la nivelul perioadei de bază se folosesc în cazul indicilor calculaţi pentru volumul fizic (care, după cum am văzut în paragrafele anterioare, se construiesc cu ajutorul sistemului de ponderare Laspeyres).

c) Serii de indici de grup cu bază în lanţ şi ponderi variabile

• Pentru caracteristica x_i:

– ponderile variabile vor fi la nivelul perioadei de bază:

$$I^{x}_{i/i-1}=\left(\frac{\sum x_i f_{i-1}}{\sum x_{i-1} f_{i-1}}\right)_{i=\overline{1,n}} \quad (7.162.)$$

$$I^{x}_{1/0}=\frac{\sum x_1 f_0}{\sum x_0 f_0};I^{x}_{2/1}=\frac{\sum x_2 f_1}{\sum x_1 f_1};\ I^{x}_{3/2}=\frac{\sum x_3 f_2}{\sum x_2 f_2};\ ...;I^{x}_{n/n-1}=\frac{\sum x_n f_{n-1}}{\sum x_{n-1} f_{n-1}} \quad (7.163.)$$

Se observă că la această serie nu se mai verifică relaţia potrivit căreia produsul indicilor cu baza în lanţ este egal cu indicele de bază fixă al întregii perioade. De fapt, această relaţie se verifică numai în cazul folosirii ponderilor constante.

– ponderi variabile la nivelul perioadei curente:

$$I^{x}_{i/i-1}=\left(\frac{\sum x_i f_i}{\sum x_{i-1} f_i}\right)_{i=\overline{1,n}} \quad (7.164.)$$

Scanned with OKEN Scanner

$$I^x_{1/0} = \frac{\sum x_1 f_1}{\sum x_0 f_1}; I^x_{2/1} = \frac{\sum x_2 f_2}{\sum x_1 f_2}; I^x_{3/2} = \frac{\sum x_3 f_3}{\sum x_2 f_3}; \ldots; I^x_{n/n-1} = \frac{\sum x_n f_n}{\sum x_{n-1} f_n} \quad (7.165.)$$

• Pentru variabila f_i:

– ponderi variabile la nivelul perioadei de bază:

$$I^f_{i/i-1} = \left(\frac{\sum x_{i-1} f_i}{\sum x_{i-1} f_{i-1}} \right)_{i=\overline{1,n}} \quad (7.166.)$$

$$I^f_{1/0} = \frac{\sum x_0 f_1}{\sum x_0 f_0}; I^f_{2/1} = \frac{\sum x_1 f_2}{\sum x_1 f_1}; I^f_{3/2} = \frac{\sum x_2 f_3}{\sum x_2 f_2}; \ldots I^f_{n/n-1} = \frac{\sum x_{n-1} f_n}{\sum x_{n-1} f_{n-1}} \quad (7.167.)$$

– ponderi variabile la nivelul perioadei curente:

$$I^f_{i/i-1} = \left(\frac{\sum x_i f_i}{\sum x_i f_{i-1}} \right)_{i=\overline{1,n}} \quad (7.168.)$$

$$I^f_{1/0} = \frac{\sum x_1 f_1}{\sum x_1 f_0}; I^f_{2/1} = \frac{\sum x_2 f_2}{\sum x_2 f_1}; I^f_{3/2} = \frac{\sum x_3 f_3}{\sum x_3 f_2}; \ldots I^f_{n/n-1} = \frac{\sum x_n f_n}{\sum x_n f_{n-1}} \quad (7.169.)$$

În practică, alegerea uneia dintre aceste variante de serii de indici se va face în funcție de conținutul indicatorului analizat și de datele disponibile.

7.8. Sisteme concrete de indici folosite în activitatea economico-socială

Datorită complexității celor mai multe fenomene economico-sociale, a numărului important de factori ce le determină, indicii statistici care ne ajută să studiem variația unor astfel de fenomene se integrează într-un sistem de indici. Datorită lui obținem o caracterizare unitară și completă a acestor fenomene.

Cele mai folosite sisteme de indicii ce caracterizează anumite laturi ale activității economico-sociale sunt:

– indicii valorii, volumului fizic și prețurilor;

– indicii productivității muncii;

– indicii salariului mediu etc.

a) Indicii valorii, volumului fizic și prețurilor

Aceștia se calculează pe baza unor date ce provin din observații totale sau parțiale.

Variabila complexă este, în acest caz, valoarea producției (a mărfurilor vândute), care se descompune în produsul a doi factori: prețul unitar al produselor

Scanned with OKEN Scanner

şi cantităţile în care vor fi vândute (produse). Pentru a studia variaţia trebuie cunoscut nivelul acestor variabile pentru două perioade – de bază şi curentă.

Aşa cum am văzut anterior, s-au folosit notaţiile: v_0, v_1, q_0, q_1, p_0, p_1.

Pentru a studia dinamica la nivelul fiecărui produs, se calculează indicii individuali ai valorii:

$$I^v = \frac{v_1}{v_0} = \frac{q_1 p_1}{q_0 p_0} = i^q \cdot i^p \qquad (7.170.)$$

unde $i^q = \frac{q_1}{q_0}$ = indicele individual al cantităţilor produse (sau vândute);

$i^p = \frac{p_1}{p_0}$ = indicele individual al preţurilor unitare ale produselor.

Pentru o viziune de ansamblu sintetică a variaţiei relative a valorii producţiei (la nivelul tuturor produselor), se calculează indicele de grup al valorii producţiei:

$$I^{\Sigma v} = \frac{\Sigma v_1}{\Sigma v_0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}$$ care este un indice în formă agregată (7.171.)

El sintetizează efectul atât al factorilor intensivi, cât şi extensivi asupra valorii producţiei. Este însă important de aflat şi măsura numerică a influenţei acestor factori asupra valorii şi de aceea se vor calcula indicii factoriali.

Folosind pentru construirea acestor indici factoriali un sistem de ponderare încrucişat, inclus în metoda substituţiei în lanţ, indicele de grup al factorului extensiv (cantitativ – cantităţile) se va construi folosind drept pondere preţul unitar la nivelul perioadei de bază (indice tip Laspeyres), iar indicele de grup al factorului intensiv (calitativ – preţul) se calculează utilizând drept ponderi – cantităţile, la nivelul perioadei curente (indice Paasche):

$$I^{\Sigma v(q)} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \qquad (7.172.)$$

$$I^{\Sigma v(p)} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \qquad (7.173.)$$

Cei doi indici factoriali şi indicele general îndeplinesc condiţiile de reversibilitate a factorilor, adică:

$$I^{\Sigma v(q)} \cdot I^{\Sigma v(p)} = I^{\Sigma v} \qquad (7.174.)$$

Pe baza indicilor calculaţi se pot determina şi modificările relative:

$$\left(I^{\Sigma v} - 1\right) \cdot 100 = R^{\Sigma v} \qquad (7.175.)$$

$$\left(I^{\Sigma v(q)} - 1\right) \cdot 100 = R^{\Sigma v(q)} \qquad (7.176.)$$

Scanned with OKEN Scanner

$$\left(I^{\Sigma v(p)} - 1\right) \cdot 100 = R^{\Sigma v(p)} \qquad (7.177.)$$

Prin diferența între numărătorul și numitorul indicilor de grup obținem modificările absolute: pe cea a valorii totale și pe cele datorate influenței celor doi factori.

$$\Delta\Sigma v = \Sigma v_1 - \Sigma v_0 = \Sigma p_1 q_1 - \Sigma p_0 q_0 \qquad (7.178.)$$

$$\Delta\Sigma v(q) = \Sigma p_0 q_1 - \Sigma p_0 q_0 \qquad (7.179.)$$

$$\Delta\Sigma v(p) = \Sigma p_1 q_1 - \Sigma p_0 q_1 \qquad (7.180.)$$

Cumulând modificările absolute datorate influenței factorilor obținem modificarea absolută totală.

$$\Delta\Sigma v = \Delta\Sigma v(q) + \Delta\Sigma v(p) \qquad (7.181.)$$

Această egalitate se verifică și la nivelul fiecărui produs.

$$\Delta v = v_1 - v_0 = q_1 p_1 - \Sigma p_0 q_0 \qquad (7.182.)$$

$$\Delta v(q) = (q_1 - q_0) p_0 \qquad (7.183.)$$

$$\Delta v(p) = (p_1 - p_0) q_1 \qquad (7.184.)$$

Uneori, însă, în evidență curentă, nu găsim toate datele de care avem nevoie pentru calculul indicilor de grup într-o formă agregată. În cazul în care se cunosc indicii individuali și valoarea producției în perioada de bază, atunci indicii de grup se pot calcula ca medii ponderate ale indicilor individuali.

Spre exemplu, indicii de grup ai valorii și ai volumului fizic se pot determina ca medii aritmetice ponderate ale indicilor individuali:

$$\left.\begin{aligned} I^{\Sigma v} &= \frac{\Sigma v_1}{\Sigma v_0} \\ i^v &= \frac{v_1}{v_0} \Rightarrow v_1 = i^v \cdot v_0 \end{aligned}\right\} \Rightarrow I^{\Sigma v} = \frac{\Sigma i^v \cdot v_0}{\Sigma v_0} = \frac{\Sigma i^v \cdot (q_0 p_0)}{\Sigma (q_0 p_0)} \qquad (7.185.)$$

$$\left.\begin{aligned} I^{\Sigma v(q)} &= \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \\ i^q &= \frac{q_1}{q_0} \Rightarrow q_1 = i^q \cdot q_0 \end{aligned}\right\} \Rightarrow I^{\Sigma v(q)} = \frac{\Sigma i^q \cdot q_0 p_0}{\Sigma q_0 p_0} = \frac{\Sigma i^q \cdot v_0}{\Sigma v_0} \qquad (7.186.)$$

$$I^{\Sigma v(p)} = \frac{I^{\Sigma v}}{I^{\Sigma v(q)}} \qquad (7.187.)$$

Dacă se cunosc indicii individuali și valoarea corespunzătoare perioadei curente, indicii de grup ai valorii și ai prețurilor se pot determina ca medie armonică ponderată a indicilor individuali:

Scanned with OKEN Scanner

$$\left.\begin{aligned} I^{\Sigma v} &= \frac{\Sigma v_1}{\Sigma v_0} \\ i^v &= \frac{v_1}{v_0} \Rightarrow v_0 = \frac{v_1}{i^v} \end{aligned}\right\} \Rightarrow I^{\Sigma v} = \frac{\Sigma v_1}{\Sigma \frac{1}{i^v} \cdot v_1} \qquad (7.188.)$$

$$\left.\begin{aligned} I^{\Sigma v(p)} &= \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} \\ i^p &= \frac{p_1}{p_0} \Rightarrow p_0 = \frac{p_1}{i^p} \end{aligned}\right\} \Rightarrow I^{\Sigma v(p)} = \frac{\Sigma v_1}{\Sigma \frac{1}{i^p} \cdot v_1} \qquad (7.189.)$$

$$I^{\Sigma v(q)} = \frac{I^{\Sigma v}}{I^{\Sigma v(p)}} \qquad (7.190.)$$

b) **Indicii productivității muncii**

În acest caz operăm cu colectivități formate din elemente omogene, de același fel, aceasta însemnând că valorile factorului cantitativ sunt direct însumabile (de exemplu, persoanele, cantități din același produs etc.)

Productivitatea este caracteristică intensivă (calitativă), derivată, întrucât ea se determină după modelul mărimilor relative de intensitate.

$$W = \frac{q}{N} - \text{la nivelul unei filiale} \qquad (7.191.)$$

unde: q = este valoarea producției la nivelul unei filiale

N = este numărul de salariați al unei filiale

Pe total, productivitatea muncii se determină ca medie aritmetică ponderată a productivităților individuale,

$$\overline{W} = \frac{\Sigma q}{\Sigma N} = \frac{\Sigma W \cdot N}{\Sigma N} = \Sigma W \cdot N^* \qquad (7.192.)$$

Mai mult, am pus în evidență doi factori ce influențează nivelul productivității medii a muncii: productivitatea individuală (N) și structura personalului $\left(N^* = \frac{N}{\Sigma N}\right)$

Indicele care redă variația relativă a productivității muncii pe total va fi calculat ca raport a două medii:

$$I^{\overline{W}} = \frac{\overline{W_1}}{\overline{W_0}} = \frac{\Sigma W_1 N_1}{\Sigma N_1} : \frac{\Sigma W_0 N_0}{\Sigma N_0} = \frac{\Sigma W_1 N_1^*}{\Sigma W_0 N_0^*} \qquad (7.193.)$$

Scanned with OKEN Scanner

Se poate pune în evidență influența celor doi factori la modificarea productivității medii a muncii:

$$I^{\overline{W}(W)} = \frac{\sum W_1 N_1^*}{\sum W_0 N_1^*}$$

$$I^{\overline{W}(g^N)} = \frac{\sum W_0 N_1^*}{\sum W_0 N_0^*}$$

$$I^{\overline{W}(W)} \cdot I^{\overline{W}(N^*)} = I^{\overline{W}(W \cdot N^*)} \qquad (7.194.)$$

Modificările absolute vor fi:

$$\Delta\overline{W} = \overline{W}_1 - \overline{W}_0 = \frac{\sum W_1 N_1}{\sum N_1} - \frac{\sum W_0 N_0}{\sum N_0} = \sum W_1 N_1^* - \sum W_0 N_0^* \qquad (7.195.)$$

$$\Delta\overline{W}(W) = \frac{\sum W_1 N_1}{\sum N_1} - \frac{\sum W_0 N_1}{\sum N_1} = \sum W_1 N_1^* - \sum W_0 N_1^* \qquad (7.196.)$$

$$\Delta\overline{W}(N^*) = \frac{\sum W_0 N_1}{\sum N_1} - \frac{\sum W_0 N_0}{\sum N_0} = \sum W_0 N_1^* - \sum W_0 N_0^* \qquad (7.197.)$$

$$\Delta\overline{W}(W) + \Delta\overline{W}(N^*) = \Delta\overline{W}(W, N^*) \qquad (7.198.)$$

c) **Indicii salariului mediu și ai fondului de salarii**

Salariul mediu este o caracteristică derivată, obținută ca raport între fondul de salarii și numărul de salariați. Vom face următoarele notații:

S = salariul obținut de salariații dintr-o grupă sau filială (compartiment) al unei societăți comerciale;

FS = fondul de salarii al grupei/compartimentului S.C.;

N = nr. de salariați dintr-o grupă/compartiment.

Pe total:

$$\overline{S} = \frac{\sum FS}{\sum N} = \frac{\sum S \cdot N}{\sum N} = \sum S \cdot N^* \qquad (7.199.)$$

Unde: $\overline{S}$ = salariul mediu la nivel total (de S.C.)

N* = structura personalului

$$N^* = \frac{N}{\sum N} \qquad (7.200.)$$

Salariul mediu total se obține ca medie aritmetică ponderată a salariilor individuale. Este influențat de doi factori: unul intensiv (S), altul extensiv (N*).

Indicele salariului mediu total, care include influența ambilor factori se calculează ca raport de medii:

Scanned with OKEN Scanner

$$I^{\bar{S}} = \frac{\overline{S_1}}{\overline{S_0}} = \frac{\sum S_1 N_1}{\sum N_1} : \frac{\sum S_0 N_0}{\sum N_0} = \frac{\sum S_1 N_1^*}{\sum S_0 N_0^*} \quad (7.201.)$$

Cei doi indici factoriali vor fi:

$$I^{\bar{S}(S)} = \frac{\sum S_1 N_1}{\sum N_1} : \frac{\sum S_0 N_1}{\sum N_1} = \frac{\sum S_1 N_1^*}{\sum S_0 N_1^*} \quad (7.202.)$$

$$I^{\bar{S}(N^*)} = \frac{\sum S_0 N_1}{\sum N_1} : \frac{\sum S_0 N_0}{\sum N_0} = \frac{\sum S_0 N_1^*}{\sum S_0 N_0^*} \quad (7.203.)$$

$$I^{\bar{S}(S,N^*)} = I^{\bar{S}(S)} \cdot I^{\bar{S}(N^*)} \quad (7.204.)$$

Modificările absolute se calculează conform relațiilor:

$$\Delta\bar{S}(S,N^*) = \overline{S_1} - \overline{S_0} = \frac{\sum S_1 N_1}{\sum N_1} - \frac{\sum S_0 N_0}{\sum N_0} = \sum S_1 N_1^* - \sum S_0 N_0^* \quad (7.205.)$$

$$\Delta\bar{S}(S) = \frac{\sum S_1 N_1}{\sum N_1} - \frac{\sum S_0 N_1}{\sum N_1} = \sum S_1 N_1^* - \sum S_0 N_1^* \quad (7.206.)$$

$$\Delta\bar{S}(N^*) = \frac{\sum S_0 N_1}{\sum N_1} - \frac{\sum S_0 N_0}{\sum N_0} = \sum S_0 N_1^* - \sum S_0 N_0^* \quad (7.207.)$$

$$\Delta\bar{S}(S,N^*) = \Delta\bar{S}(S) + \Delta\bar{S}(N^*) \quad (7.208.)$$

Pentru dinamica şi modificarea absolută a fondului de salarii, se procedează astfel:

I

$$FS = S \cdot N \quad (7.209.)$$

$$\sum FS = \sum S \cdot N = \bar{S} \cdot \sum N \quad (7.210.)$$

$$I^{\Sigma FS} = \frac{\sum FS_1}{\sum FS_0} = \frac{\sum S_1 N_1}{\sum S_0 N_0} \begin{cases} I^{\Sigma FS(S)} = \dfrac{\sum S_1 N_1}{\sum S_0 N_1} \\ I^{\Sigma FS(N)} = \dfrac{\sum S_0 N_1}{\sum S_0 N_0} \end{cases} \quad (7.211.)$$

$$\Delta\Sigma FS = \Sigma FS_1 - \Sigma FS_0 = \Sigma S_1 N_1 - \Sigma S_0 N_0 \quad (7.212.)$$

$$\begin{cases} \Delta\Sigma FS(S) = \Sigma S_1 N_1 - \Sigma S_0 N_1 \\ \Delta\Sigma FS(N) = \Sigma S_0 N_1 - \Sigma S_0 N_0 \end{cases} \quad (7.213.)$$

Scanned with OKEN Scanner

II

$$I^{\Sigma FS} = \frac{\overline{S_1}\Sigma N_1}{\overline{S_0}\Sigma N_0} \begin{cases} I^{\Sigma FS(\overline{S})} = \dfrac{\overline{S_1}\Sigma N_1}{\overline{S_0}\Sigma N_1} = I^{\overline{S}} & (7.214.) \\ I^{\Sigma FS(\Sigma N)} = \dfrac{\overline{S_0}\Sigma N_1}{\overline{S_0}\Sigma N_0} = I^{\Sigma N} & (7.215.) \end{cases}$$

$$\Delta\Sigma FS = \overline{S_1}\Sigma N_1 - \overline{S_0}\Sigma N_0 \begin{cases} \Delta\Sigma FS(\overline{S}) = (\overline{S_1} - \overline{S_0})\Sigma N_1 & (7.216.) \\ \Delta\Sigma FS(\Sigma N) = \overline{S_0}(\Sigma N_1 - \Sigma N_0) & (7.217.) \end{cases}$$

III

$$I^{\Sigma FS} = \frac{(\Sigma S_1 N_1^*)\Sigma N_1}{(\Sigma S_0 N_0^*)\Sigma N_0}$$

$$I^{\Sigma FS(\Sigma N)} = \frac{(\Sigma S_0 N_0^*)\Sigma N_1}{(\Sigma S_0 N_0^*)\Sigma N_0} \qquad (7.218.)$$

$$I^{\Sigma FS(N^*)} = \frac{(\Sigma S_0 N_1^*)\Sigma N_1}{(\Sigma S_0 N_0^*)\Sigma N_1} \qquad (7.219.)$$

$$I^{\Sigma FS(S)} = \frac{(\Sigma S_1 N_1^*)\Sigma N_1}{(\Sigma S_0 N_1^*)\Sigma N_1} \qquad (7.220.)$$

$$\Delta\Sigma FS(S, N^*, \Sigma N) = (\Sigma S_1 N_1^*)\Sigma N_1 - (\Sigma S_0 N_0^*)\Sigma N_0 \qquad (7.221.)$$

$$-\Delta\Sigma FS(\Sigma N) = (\Sigma S_0 N_0^*)\Sigma N_1 - (\Sigma S_0 N_0^*)\Sigma N_0 \qquad (7.222.)$$

$$-\Delta\Sigma FS(N^*) = (\Sigma S_0 N_1^*)\Sigma N_1 - (\Sigma S_0 N_0^*)\Sigma N_1 \qquad (7.223.)$$

$$-\Delta\Sigma FS(S) = (\Sigma S_1 N_1^*)\Sigma N_1 - (\Sigma S_0 N_1^*)\Sigma N_1 \qquad (7.224.)$$

S-au pus în evidență trei modalități de descompunere a fondurilor de salarii, astfel:

I. în funcție de doi factori:
- salariul individual;
- numărul de salariați la nivelul unei grupe/compartiment.

II. în funcție de doi factori:
- salariul mediu pe total;
- numărul total de salariați.

III. în funcție de trei factori:
- salariul individual;
- structura salariaților;
- numărul total de salariați.

Scanned with OKEN Scanner

În funcție de aceste descompuneri, s-au calculat indicii totali și factoriali, ca și modificările absolute totale și ale influențelor factorilor.

Pentru determinarea indicilor factoriali s-a folosit metoda substituției în lanț.

7.9. Indici teritoriali

Indicii teritoriali măsoară variația medie relativă a unui fenomen în plan teritorial. Indicele teritorial se calculează ca raport a două niveluri diferite ale aceluiași fenomen, înregistrat în unități de spațiu diferite. El este, de fapt, o mărime relativă de coordonare.

Prin analiza variației în profil teritorial se pot lua măsuri în vederea atenuării și eliminării discrepanțelor între regiuni sau alte unități teritoriale.

Indicii teritoriali se construiesc pe aceleași principii ca și indicii în general, ținându-se seama de unele particularități ale analizei în plan teritorial:

– stabilirea bazei de raportare;
– stabilirea ponderilor folosite;
– alegerea procedeului de calcul;
– alegerea metodei de defalcare a variației pe factori de influență.

Baza de raportare se alege în urma unei analize atente și complexe asupra fenomenului a cărui variație o studiem. Unitatea teritorială aleasă ca bază de raportare trebuie să prezinte un nivel comun obișnuit al fenomenului, nu unul de excepție. Unitatea teritorială – baza de comparație nu se poate menține neschimbată o perioadă mare de timp, datorită modificărilor care apar în activitatea economico-socială în profil teritorial. Alegerea unității teritoriale – bază de comparație se va face și după o anumită logică, după scopul, obiectivul urmărit în cercetarea statistică.

Indicii teritoriali se calculează în două ipostaze:

a) ca indici individuali – la nivelul unor unități teritoriale simple;
b) ca indici de grup – la nivelul unor unități complexe.

a) **indici individuali**

Fie i, j – două unități teritoriale.

Indicii teritoriali individuali ce se pot construi sunt:

$$i^{y}_{i/j} = \frac{y_i}{y_j} \quad \text{dacă } j \text{ este unitatea teritorială aleasă ca bază de comparație} \qquad (7.225.)$$

Scanned with OKEN Scanner

$i^{y}_{j/i} = \frac{y_j}{y_i}$ dacă *i* este unitatea teritorială aleasă ca bază de comparație (7.226.)

Produsul celor doi indici este 1, ceea ce înseamnă că ei îndeplinesc principiul reversibilității în spațiu. Totuşi, în general, nu se utilizează ambii indici, alegându-se ca bază de raportare acea unitate (*i* sau *j*) care prezintă interes pentru cercetători.

b) **indici de grup** – în funcție de natura variabilei, a fenomenului analizat, pentru care se calculează, indicii de grup se determină prin mai multe metode:

– dacă valorile variabilei cantitative sunt însumabile direct, atunci indicii vor avea expresia:

$$I^{\Sigma f}_{i/j} = \frac{\Sigma f_i}{\Sigma f_j} \text{ sau } I^{\Sigma f}_{j/i} = \frac{\Sigma f_j}{\Sigma f_i} \qquad (7.227.)$$

cu $$I^{\Sigma f}_{i/j} \cdot I^{\Sigma f}_{j/i} = 1 \qquad (7.228.)$$

În acest caz, variabila calitativă se determină ca o mărime medie:

$$\bar{x} = \frac{\Sigma y}{\Sigma f} \qquad (7.229.)$$

Indicele care reflectă variația în profil teritorial a variabilei calitative se obține conform relației:

$$I^{\bar{x}}_{i/j} = \frac{\bar{x}_i}{\bar{x}_j} = \frac{\Sigma y_i}{\Sigma f_i} : \frac{\Sigma y_j}{\Sigma f_j} = \frac{I^{\Sigma y}_{i/j}}{I^{\Sigma f}_{i/j}} \qquad (7.230.)$$

$$I^{\Sigma y}_{i/j} = I^{\bar{x}}_{i/j} \cdot I^{\Sigma f}_{i/j} \qquad (7.231.)$$

$$\Sigma y = \bar{x} \cdot \Sigma f \qquad (7.232.)$$

Aşadar, şi în cazul indicilor teritoriali, se verifică relația conform căreia produsul indicilor factoriali este egal cu indicele variabilei complexe.

– dacă valorile variabilei factorului cantitativ nu sunt însumabile direct, în construirea indicelui de grup se impune folosirea unui comăsurător (pondere), reprezentat de celălalt factor (calitativ). Ponderea poate fi valoarea factorului calitativ la nivelul unității *i* sau la nivelul unității *j*. În funcție de aceasta, indicele de grup al factorului cantitativ este:

$$I^{f}_{ij} = \frac{\Sigma x_i f_i}{\Sigma x_i f_j} \qquad (7.233.)$$

$$I^{f}_{ij} = \frac{\Sigma x_j f_i}{\Sigma x_j f_j} \qquad (7.234.)$$

Scanned with OKEN Scanner

Analog, pentru construirea indicelui de grup al factorului calitativ, se va folosi drept pondere factorul cantitativ la nivelul unității *i* sau *j*.

$$I_{i/j}^{x} = \frac{\Sigma x_i f_i}{\Sigma x_j f_i} \qquad (7.235.)$$

$$I_{i/j}^{x} = \frac{\Sigma x_i f_j}{\Sigma x_j f_j} \qquad (7.236.)$$

În cazul utilizării ponderilor, nu se mai verifică testul reversibilității factorilor. Utilizarea ponderării încrucișate nu se utilizează în cazul comparațiilor teritoriale, astfel că în practică se aplică un indice ce seamănă cu indicele Fisher, construit ca medie geometrică a indicilor calculați în cele două variante de ponderare:

$$I_{i/j}^{f} = \sqrt{\frac{\Sigma x_i f_i}{\Sigma x_i f_j} \cdot \frac{\Sigma x_j f_i}{\Sigma x_j f_j}} \qquad (7.237.)$$

$$I_{i/j}^{x} = \sqrt{\frac{\Sigma x_i f_i}{\Sigma x_j f_i} \cdot \frac{\Sigma x_i f_j}{\Sigma x_j f_j}} \qquad (7.238.)$$

Scanned with OKEN Scanner

ANEXE

Anexa 1

Funcția Gauss-Laplace

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{u^2}{2}}du$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0,0	0.0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0.2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Anexa 2

Tabel cu valorile repartiției Student în funcție de probabilitatea $P(t \le t_\alpha)$ și numărul f al gradelor de libertate

	Nivel de semnificație pentru testul bilateral								
f \ α	0,50	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001
1	1,000	3,078	6,314	12,706	31,821	63,657	318,309	636,618	6366,198
2	0,816	1,886	2,920	4,303	6,965	9,925	22,327	31,598	99,992
3	0,765	1,638	2,353	3,482	4,541	5,841	10,214	12,924	28,000
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610	15,544
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869	11,178
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959	9,082
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7,885
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041	7,120
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781	6,594
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587	6,211
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437	5,921
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318	5,694
13	0,694	1,350	1,771	2,160	2,650	3,102	3,852	4,221	5,513
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140	5,363
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073	5,239
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015	5,134
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965	5,014
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922	4,966
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883	4,897
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850	4,837
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819	4,784
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792	4,736
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767	4,693
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745	4,654
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725	4,619
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707	4,587
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690	4,558
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674	4,530
29	0,683	1,312	1,699	2,045	2,462	2,756	3,396	3,659	4,506
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646	4,482
35	0,682	1,306	1,690	2,030	2,438	2,724	3,340	3,591	4,389
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551	4,321
45	0,680	1,301	1,679	2,014	2,412	2,690	3,281	3,520	4,269
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496	4,288
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460	4,169
70	0,678	1,294	1,667	1,994	2,381	2,648	3,211	3,435	4,127
80	0,678	1,292	1,664	1,990	2,374	2,639	3,195	3,416	4,096
90	0,677	1,291	1,662	1,987	2,368	2,632	3,183	3,402	4,072
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390	4,053
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373	4,025
200	0,676	1,286	1,653	1,972	2,345	2,601	3,131	3,340	3,970
500	0,675	1,283	1,648	1,965	2,334	2,586	3,107	3,310	3,922
1000	0,675	1,282	1,646	1,962	2,330	2,581	3,098	3,300	3,906
∞	0,675	1,282	1,645	1,960	2,326	2,576	3,090	3,290	3,891
f \ α	0,25	0,10	0,05	0,025	0,01	0,005	0,001	0,0005	0,00005
	Nivel de semnificație pentru testul unilateral								

Anexa 3

Repartiția F (Fisher Snedecor)

Valorile funcției F ($F = s_1^2 / s_2^2$) pentru l_1, l_2 grade de libertate și $\alpha = 0,05$, $\alpha = 0,01$ nivel de semnificație

	$l_1 = 1$		$l_1 = 2$		$l_1 = 3$		$l_1 = 4$		$l_1 = 5$	
l_2	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
1	161,40	4052,00	199,50	4999,00	215,70	5403,00	224,60	5625,00	230,20	5764,00
2	18,51	98,49	19,00	99,00	19,16	99,17	19,25	99,25	19,30	99,30
3	10,13	34,12	9,55	30,81	9,28	29,46	9,12	28,71	9,01	28,24
4	7,71	21,20	6,94	18,00	6,59	16,69	6,39	15,98	6,26	15,52
5	6,61	16,26	5,79	13,27	5,41	12,06	5,19	11,39	5,05	10,97
6	5,99	13,74	5,14	10,91	4,76	9,78	4,53	9,15	4,39	8,75
7	5,59	12,25	4,74	9,55	4,35	8,45	4,12	7,85	3,97	7,45
8	5,32	11,26	4,46	8,65	4,07	7,59	3,84	7,01	3,69	6,63
9	5,12	10,56	4,26	8,02	3,86	6,99	3,63	6,42	3,48	6,06
10	4,96	10,04	4,10	7,56	3,71	6,55	3,48	5,99	3,33	5,64
11	4,84	9,65	3,98	7,20	3,59	6,22	3,36	5,67	3,20	5,32
12	4,75	9,33	3,88	6,93	3,49	5,95	3,26	5,41	3,11	5,06
13	4,67	9,07	3,80	6,70	3,41	5,74	3,18	5,20	3,02	4,86
14	4,60	8,86	3,74	6,51	3,34	5,56	3,11	5,03	2,96	4,69
15	4,54	8,68	3,68	6,36	3,29	5,42	3,06	4,89	2,90	4,56
16	4,49	8,53	3,63	6,23	3,24	5,29	3,01	4,77	2,85	4,44
17	4,45	8,40	3,59	6,11	3,20	5,18	2,96	4,67	2,81	4,34
18	4,41	8,28	3,55	6,01	3,16	5,09	2,93	4,58	2,77	4,25
19	4,38	8,18	3,52	5,93	3,13	5,01	2,90	4,50	2,74	4,17
20	4,35	8,10	3,49	5,85	3,10	4,94	2,87	4,43	2,71	4,10
21	4,32	8,02	3,47	5,78	3,07	4,87	2,84	4,37	2,68	4,04
	$l_1 = 1$		$l_1 = 2$		$l_1 = 3$		$l_1 = 4$		$l_1 = 5$	
l_2	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
22	4,30	7,94	3,44	5,72	3,05	4,82	2,82	4,31	2,66	3,99
23	4,28	7,88	3,42	5,66	3,03	4,76	2,80	4,26	2,64	3,94
24	4,26	7,82	3,40	5,61	3,01	4,72	2,78	4,22	2,62	3,90
25	4,24	7,77	3,38	5,57	2,99	4,68	2,76	4,18	2,60	3,86

Scanned with OKEN Scanner

26	4,22	7,72	3,37	5,53	2,98	4,64	2,74	4,14	2,59	3,82
27	4,21	7,68	3,35	5,49	2,96	4,60	2,73	4,11	2,57	3,78
28	4,20	7,64	3,34	5,45	2,95	4,57	2,71	4,07	2,56	3,75
29	4,18	7,60	3,33	5,42	2,93	4,54	2,70	4,04	2,54	3,73
30	4,17	7,56	3,32	5,39	2,92	4,51	2,69	4,02	2,53	3,70
40	4,08	7,31	3,23	5,18	2,84	4,31	2,61	3,83	2,45	3,51
60	4,00	7,08	3,15	4,98	2,76	4,13	2,52	3,65	2,37	3,34
120	3,92	6,85	3,07	4,79	2,68	3,95	2,45	3,48	2,29	3,17
∞	3,84	6,64	2,99	4,60	2,60	3,78	2,37	3,32	2,21	3,02

Scanned with OKEN Scanner

Anexa 4

Valorile variabilei χ^2_α în funcție de probabilitatea $\alpha = P(\chi^2 > \chi^2_\alpha)$

α / l	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	190,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Scanned with OKEN Scanner

Anexa 5

Tabel cu numere aleatorii

Rân-dul	Coloana 1	2	3	4	5	6	7	8	9	10
1	87 35	67 44	51 49	18 98	97 84	75 22	53 29	10 52	26 87	54 92
2	25 52	29 67	35 99	48 88	40 08	63 68	82 39	38 47	91 39	11 00
3	87 17	83 31	25 59	87 48	25 80	24 08	81 45	21 32	90 08	44 31
4	05 04	40 35	72 95	48 56	77 57	63 19	80 16	48 52	06 47	64 98
5	81 16	09 21	91 71	29 76	54 01	53 17	30 67	62 95	56 58	10 91
6	54 85	78 88	57 91	11 69	10 22	71 87	24 92	52 64	42 82	78 95
7	44 78	19 18	35 40	27 66	89 72	21 17	71 69	95 17	97 17	62 60
8	97 20	98 97	37 33	93 75	18 88	35 85	46 05	07 20	08 17	66 24
9	98 77	57 51	40 41	76 21	18 54	60 61	79 13	94 57	50 73	89 68
10	78 12	77 30	83 30	59 28	73 33	47 07	60 07	45 38	82 10	73 19
11	41 19	70 62	43 46	06 13	22 38	31 18	61 60	07 14	49 16	28 16
12	70 64	30 55	67 46	95 79	63 66	82 56	67 10	76 77	03 22	42 18
13	06 56	09 89	68 87	79 19	35 94	66 18	17 91	72 81	72 77	92 39
14	29 46	18 28	08 88	48 56	49 44	67 82	72 67	28 83	10 26	58 13
15	42 14	55 51	72 95	29 25	15 18	25 68	48 92	87 16	78 43	17 47
16	33 75	87 15	15 23	13 79	62 73	76 69	09 77	82 65	72 47	59 56
17	09 80	99 61	98 08	34 11	88 79	08 32	46 78	35 58	44 16	12 23
18	98 31	57 50	85 80	53 39	05 92	54 42	29 01	35 23	09 81	96 64
19	51 70	52 55	83 12	95 02	79 11	49 79	87 95	98 48	88 08	64 77
20	27 83	61 07	49 05	46 20	35 78	31 34	42 50	68 11	42 14	29 77
21	78 84	69 15	64 42	92 39	36 08	56 39	35 02	92 78	46 63	82 98
22	22 12	89 66	49 09	99 10	62 53	19 31	81 83	50 43	37 42	10 00
23	69 41	59 54	82 72	44 66	64 03	76 59	12 12	41 56	34 90	26 06
24	54 99	46 54	51 38	59 07	64 21	81 17	88 47	23 05	63 43	08 67
25	99 91	82 79	92 62	44 24	01 34	45 16	33 56	17 78	42 86	70 94
26	54 96	58 70	68 38	48 37	08 25	51 28	42 09	45 96	78 84	23 59
27	88 37	31 01	58 97	65 10	96 97	32 75	65 73	06 01	60 19	68 10
28	49 40	84 90	19 91	75 41	55 63	98 25	01 69	29 57	72 77	75 95
29	21 18	87 99	98 48	38 12	64 18	78 50	43 97	04 07	85 18	98 87
30	06 31	99 51	27 13	43 21	07 55	90 35	29 84	08 70	66 03	18 96
31	58 56	75 14	10 32	35 15	47 05	38 54	92 08	08 72	79 66	65 20
32	83 00	61 74	76 68	58 67	31 58	77 93	92 44	94 16	01 29	32 19
33	51 45	05 35	16 73	80 43	44 64	19 14	52 22	81 62	79 68	13 37
34	90 55	27 93	33 06	76 42	16 08	16 81	86 15	09 79	30 86	74 77
35	68 65	05 79	82 57	78 22	37 41	82 44	07 43	98 89	78 23	96 04
36	51 84	34 98	03 38	87 12	79 88	57 88	46 81	62 56	94 91	06 59
37	85 23	92 01	84 05	83 64	12 95	51 44	60 64	68 86	53 05	26 16
38	31 95	14 27	80 20	36 96	81 84	88 55	50 47	22 39	70 56	66 41
39	91 55	02 31	42 31	54 72	45 32	43 76	89 95	74 04	79 33	46 91
40	46 80	10 25	92 61	83 83	58 58	71 46	29 95	72 38	89 05	26 05
41	07 78	66 31	63 37	30 38	95 57	84 99	85 32	91 28	06 85	67 13
42	48 69	61 99	15 93	98 43	38 84	46 49	84 89	44 51	75 42	45 17
43	70 69	68 47	02 10	60 99	02 84	41 36	81 65	44 08	49 93	49 62
44	98 63	62 58	95 62	61 53	28 49	20 20	21 49	45 35	01 57	45 64
45	76 50	34 97	87 14	65 07	59 45	02 56	89 50	85 03	42 65	07 35
46	23 47	83 35	73 08	83 52	47 19	79 85	26 07	53 80	00 33	18 07
47	18 38	27 88	50 09	58 11	96 83	05 86	49 61	35 03	46 63	88 48
48	93 93	51 63	96 33	82 81	58 02	19 36	55 80	74 84	14 63	79 22
49	65 23	13 22	0110	20 01	11 15	36 43	49 49	19 24	77 70	95 08
50	77 19	27 34	65 55	62 52	47 73	09 34	58 38	28 28	18 15	82 65

Scanned with OKEN Scanner

Anexa 5 (continuare)

Rân-dul	Coloana									
	1	2	3	4	5	6	7	8	9	10
51	61 39	57 28	82 11	49 79	26 68	85 55	27 69	41 27	70 74	51 72
52	20 79	22 91	99 39	15 98	39 00	87 41	31 42	37 79	94 48	90 37
53	96 28	37 52	96 74	84 48	85 77	65 63	36 72	11 78	83 00	01 39
54	35 25	98 71	48 26	14 39	73 31	81 15	76 22	17 76	03 72	32 70
55	65 29	37 18	09 40	08 51	59 45	46 94	29 57	71 74	74 19	16 83
56	30 06	86 39	75 15	85 61	83 12	60 77	88 14	67 62	50 00	59 77
57	62 63	22 21	37 17	88 28	54 66	45 52	75 40	99 32	81 19	22 04
58	44 86	78 61	36 08	79 10	56 33	15 56	66 39	74 69	73 49	36 74
59	70 30	42 32	63 10	22 08	32 39	43 53	02 33	45 90	62 92	50 23
60	53 31	94 17	71 65	36 71	69 62	96 13	24 22	91 24	36 89	98 05
61	01 19	00 22	89 97	50 76	89 94	09 25	09 40	22 55	00 11	94 82
62	59 53	30 97	09 82	69 01	51 35	69 28	68 18	68 62	34 37	36 16
63	12 64	27 42	66 42	47 13	10 10	52 61	34 19	91 68	00 50	19 99
64	83 91	19 54	89 49	54 79	30 64	16 09	07 34	69 84	09 30	11 44
65	16 84	22 33	12 54	06 08	44 90	77 47	41 03	68 80	32 94	60 48
66	64 20	55 02	45 72	18 37	53 54	01 07	13 90	41 61	99 14	79 21
67	83 21	49 79	27 82	84 32	91 86	21 35	30 65	29 07	42 09	69 16
68	58 34	75 23	57 30	02 73	03 17	06 18	71 83	98 61	47 77	12 32
69	49 74	56 30	49 30	04 11	31 75	07 02	03 56	36 78	66 59	03 66
70	14 36	53 40	17 30	00 12	32 71	07 35	92 67	32 69	77 14	47 44
71	56 77	08 01	57 35	91 01	34 17	82 26	17 74	73 63	28 82	81 52
72	62 47	85 70	90 70	96 78	94 75	63 83	92 83	15 17	22 98	66 54
73	62 33	40 58	01 57	14 36	29 50	42 99	77 10	90 09	65 93	78 59
74	00 51	35 07	12 40	90 41	69 86	17 61	75 05	63 77	50 23	06 87
75	55 30	12 69	80 94	51 03	79 62	86 41	57 05	71 98	13 08	82 17
76	15 05	72 68	21 08	61 57	62 54	74 83	24 07	86 09	21 14	40 95
77	24 56	81 69	20 97	25 19	59 55	80 77	81 34	56 35	38 42	33 01
78	32 97	31 52	09 41	18 77	88 79	38 43	07 74	33 39	54 26	02 66
79	55 11	78 50	47 47	17 94	44 98	16 93	09 55	53 73	54 00	52 26
80	48 11	03 79	08 84	00 32	26 94	50 67	98 35	98 68	68 04	86 09
81	74 24	53 19	95 66	61 98	45 46	89 64	44 73	56 57	91 52	39 56
82	62 35	99 91	01 37	76 51	83 10	63 92	61 96	48 59	39 75	60 18
83	90 14	34 82	90 53	25 80	96 50	63 41	52 21	41 51	15 10	44 53
84	35 19	70 16	35 40	68 61	42 32	39 52	93 16	18 67	05 37	21 44
85	10 34	98 89	60 50	54 77	33 50	03 89	09 02	83 11	46 82	26 31
86	95 88	45 42	65 85	77 56	59 71	96 40	49 10	70 55	02 98	64 44
87	93 90	00 60	22 99	86 12	00 21	48 74	74 90	99 82	75 08	93 39
88	68 22	99 19	65 37	71 50	69 31	56 45	59 63	50 92	63 52	02 35
89	78 44	32 84	79 93	84 63	86 57	13 70	87 01	26 91	55 91	58 67
90	03 58	35 81	62 40	53 21	63 24	83 66	37 48	94 67	11 66	21 73
91	28 91	01 98	65 56	38 27	09 45	15 13	41 60	38 99	89 78	94 29
92	00 40	05 09	71 05	35 60	25 77	19 17	30 54	38 37	82 14	86 42
93	85 61	58 86	74 13	13 69	23 03	61 36	52 86	27 92	45 59	61 63
94	06 35	53 06	68 30	23 00	65 54	30 13	69 65	61 28	61 28	12 73
95	39 69	61 11	02 81	98 55	93 20	65 13	75 06	27 21	27 21	45 71
96	12 26	23 24	42 18	34 64	08 47	39 64	96 42	49 20	28 10	91 80
97	76 93	55 27	27 17	55 98	17 07	77 29	29 88	32 20	87 18	90 95
98	52 08	40 32	40 33	67 70	25 18	45 79	33 53	29 99	88 35	97 59
99	91 49	00 88	14 76	96 02	50 36	93 96	07 29	97 18	55 79	84 63
100	16 78	05 19	94 15	12 18	25 87	93 58	30 60	49 28	05 57	47 17
101	02 22	85 19	48 74	55 24	89 69	15 53	00 20	88 48	95 08	00 47
102	85 76	34 51	40 44	62 93	65 99	72 64	09 34	01 13	09 74	90 65
103	00 88	96 79	38 24	77 00	70 91	47 43	43 82	71 67	49 90	37 00

Scanned with OKEN Scanner

Anexa 5 (continuare)

Rân-dul	Coloana									
	1	2	3	4	5	6	7	8	9	10
104	64 29	81 85	50 47	36 50	91 19	09 15	98 75	60 58	33 15	51 44
105	94 03	80 04	21 49	54 91	77 85	00 45	68 23	12 94	23 44	36 88
106	42 28	52 73	06 41	37 47	47 31	52 99	89 92	22 81	86 55	99 09
107	09 27	52 72	49 11	30 93	33 29	54 17	54 48	47 42	04 79	18 64
108	54 68	64 07	85 32	05 96	54 79	57 43	96 97	30 72	12 19	41 70
109	25 04	92 29	71 11	64 10	42 23	23 67	01 19	20 58	35 93	39 46
110	28 58	32 91	95 28	42 36	98 59	66 32	15 51	46 63	57 10	83 55
111	64 35	04 62	24 87	44 85	45 68	41 66	19 17	13 09	63 37	15 33
112	61 05	55 88	25 01	15 77	12 90	69 34	36 93	52 39	36 23	59 73
113	98 93	18 93	86 98	99 04	75 28	30 05	12 09	57 35	90 15	98 07
114	61 89	35 47	16 32	20 16	78 52	82 37	26 33	67 42	11 93	35 61
115	94 40	82 18	06 61	54 67	03 66	76 82	90 31	71 90	39 27	97 85
116	54 38	58 65	27 70	93 57	59 00	63 56	18 79	85 52	21 03	03 16
117	63 70	89 23	76 46	97 70	00 62	15 35	97 42	47 54	60 60	78 12
118	61 58	65 62	81 29	69 71	95 53	53 69	20 95	66 60	50 70	22 97
119	51 68	98 15	05 64	43 32	74 03	44 63	52 38	67 59	56 69	11 14
120	59 25	41 48	64 79	62 26	87 86	94 30	43 54	26 98	61 38	53 44
121	85 00	02 24	67 85	88 10	34 01	54 53	23 77	33 11	19 68	13 50
122	01 46	87 56	19 19	19 43	70 25	24 29	48 22	44 81	35 40	33 23
123	42 41	25 10	87 27	77 28	05 90	73 03	95 46	88 82	25 02	05 00
124	03 57	14 03	17 80	47 85	94 49	89 55	10 27	19 50	20 37	02 71
125	18 95	93 40	45 53	04 57	17 03	34 54	83 91	69 02	90 72	98 45
126	74 11	04 66	68 52	70 11	97 01	55 36	63 49	42 68	82 15	48 64
127	31 54	98 82	61 64	40 50	42 48	96 84	82 42	55 15	72 34	90 96
128	85 51	93 55	89 63	47 92	88 42	00 08	21 52	27 28	77 48	02 42
129	19 95	97 55	27 91	15 20	96 25	48 75	49 95	88 68	36 09	66 17
130	75 74	55 98	33 02	36 99	11 84	07 71	40 65	95 54	01 90	44 32

Anexa 6

Valorile coeficienților V_p $\left(V_p = \frac{x_n - \bar{x}}{s}\right)$

n \ P	0,95	0,975	0,99
3	1,412	1,414	1,414
5	1,869	1,917	1,955
7	2,093	2,182	2,265
9	2,237	2,349	2,464
10	2,294	2,414	2,540
11	2,343	2,470	2,606
12	2,387	2,519	2,663
13	2,426	2,562	2,714
14	2,461	2.602	2,759
15	2,493	2,638	2,800
16	2,523	2,670	2,837
17	2,551	2,701	2,871
18	2,577	2,728	2,903
19	2,600	2,754	2,932
20	2,623	2,778	2,959
21	2,644	2,801	2,984
22	2,664,,	2,823	3,008
23	2,683	2,842	3,030
24	2,701	2,862	3,051
25	2,717	2,880	3,071

Scanned with OKEN Scanner

Anexa 7

Valorile funcției de repartiție a amplitudinii împrăștierii

$$P = P\left(\frac{x_{(n)} - x_{(1)}}{\sigma} \le t\right)$$

t	n				
	5	6	7	8	9
4,00	0,9623	0,9469	0,9300	0,9120	0,8929
4,10	0,9693	0,9566	0,9425	0,9273	0,9112
4,20	0,9752	0,9647	0,9530	0,9404	0,9269
4,25	0,9777	0,9682	0,9576	0,9461	0,9338
4,35	0,9821	0,9744	0,9657	0,9562	0,9460
4,45	0,9857	0,9795	0,9724	0,9647	0,9563
4,50	0,9873	0,9817	0,9754	0,9684	0,9608
4,55	0,9887	0,9837	0,9780	0,9917	0,9649
4,60	0,9899	0,9855	0,9804	0,9747	0,9686
4,65	0,9911	0,9871	0,9825	0,9775	0,9719
4,70	0,9921	0,9885	0,9845	0,9799	0,9750
4,75	0,9930	0,9898	0,9862	0,9822	0,9777
4,80	0,9938	0,9910	0,9878	0,9842	0,9802
4,85	0,9945	0,9920	0,9892	0,9860	0,9824
4,90	0,9952	0,9930	0,9904	0,9876	0,9844
4,95	0,9958	0,9938	0,9916	0,9890	0,9862
5,00	0,9963	0,9946	0,9926	0,9903	0,9876
5,05	0,9967	0,9952	0,9935	0,9915	0,9893
5,10	0,9971	0,9958	0,9942	0,9925	0,9906
5,15	0,9975	0,9963	0,9950	0,9934	0,9917
5,20	0,9978	0,9968	0,9956	0,9942	0,9927
5,25	0,9981	0,9972	0,9961	0,9949	0,9936
5,30	0,9983	0,9975	0,9966	0,9956	0,9944
5,35	0,9985	0,9979	0,9971	0,9961	0,9951
5,40	0,9987	0,9980	0,9974	0,9966	0,9957
5,45	0,9989	0,9984	0,9978	0,9971	0,9963
5,50	0,9994	0,9986	0,9981	0,9975	0,9968
5,70	0,9995	0,9992	0,9989	0,9986	0,9989
6,00	0,9998	0,9997	0,9996	0,9994	0,9993

Scanned with OKEN Scanner

1530

BIBLIOGRAFIE

Anghelache C., Isaic-Maniu Al., Mitruţ C., Voineagu V., *Sistemul Conturilor Naţionale*, Editura Economică, Bucuresti, 2005.

Biji E., Lilea E., Vătui M., Matache S., Şerban D., *Statistica managerială a agentului economic din agricultură*, Editura Ceres, Bucureşti, 1998.

Curvin J., Slater R., *Quantitative Methods for Business Decisions*, Ed. Chapman & Hall, Londra, 1995.

Fourastie J., Grais B., *Les Indices Statistiques*, Masson, Paris, 1984.

Ghiţă S., *Statistica resurselor de muncă*, Editura Meteor Press, Bucureşti, 2005.

Ghiţă S., *Using Statistical Methods for Comparative Regional Analysis of the Electronic Commerce Activity*, în „Economic Computation and Economic Cybernetics Studies and Research", nr. 1-2/2006, vol. 40, p. 187-194.

Ghiţă S., *Aplicarea modelelor de regresie liniară multiplă în studierea fenomenului de şomaj*, în „Studii şi cercetări de calcul economic şi cibernetică economică", nr. 2/2006, pp. 97-104.

Ghiţă S. *Utilizarea analizei dispersionale în studiul statistic al ocupării forţei de muncă*, în „Revista Română de Statistică" nr. 5/2005, INS, Bucureşti, pp. 51-67.

Isaic-Maniu Al., Mitruţ C., Voineagu V., *Statistică*, Editura Universitară, Bucureşti, 2003.

Johnson R., *Elementary Statistics*, 4th Edition, PWS-KENT Publishing Company, Boston.

Keller G., Warrack B., *Essentials of Business Statistics – A Systematic Approach*, Wadsworth Publishing Company, Belmont, California, 1991.

Lilea E., Goschin Z., Vătui M., Boldeanu D., *Statistica*, Editura ASE, Bucureşti, 2001.

Marshall A., *Principles of Economics*, Eighth Edition, McMillan Press, London, 1992.

Pecican E. Şt., *Econometrie*, Editura ALL, Bucureşti, 1994.

Pecican E. Şt., *Macroeconometrie*, Editura Economică, Bucureşti, 1996.

Rumeau-Rouqette C., Breart G., Padieu R., *Methodes en Epidemiologie*, Paris, Flammarion Medecine-Sciences, 1983.

Runyon R., Haber A., Pittenger D., Coleman K., *Fundamentals of Behavioral Statistics*, Eight Edition, Mc. Graw Hill, USA, 1996.

Samuelson P., *Economics*, McGrow Hill Book Company, International Edition, 1989.

Ţiţan E., Ghiţă S., Trandaş C., *Bazele statisticii* – ediţie revizuită şi completată, Editura Meteor Press, Bucureşti, 2002.

Ţiţan E., Ghiţă S., Trandaş C., *Bazele statisticii*, Editura Meteor Press, Bucureşti, 2001.

Ţiţan E., Ghiţă S., Trandaş C., *Statistică aplicată*, Editura Meteor Press, Bucureşti, 2003.

Ţiţan E., *Statistică. Teorie şi aplicaţii în sectorul terţiar*, Editura Meteor Press, Bucureşti, 2002.

Voineagu V., Furtună F.T., Voineagu M., Ştefănescu C., *Analiza factorială a fenomenelor social-economice în profil regional*, Editura Aramis Print, Bucureşti, 2002.

Voineagu V., Lilea E., Goschin Z., Vătui M., *Statistica economică*, Tribuna economică, Bucureşti, 2003.

Voineagu V., Lilea E., Vătui M., *Statistică economică: teorie şi aplicaţii*, Tribuna economică, Bucureşti, 2002.

Voineagu V., Lilea E., Vătui M., *Statistică economică*, Editura ASE, Bucureşti, 2000.

Voineagu V., Isaic-Maniu Al., Mitruţ C., *Statistică*, Editura Cison, Bucureşti, 2002.

Voineagu V., *Statistica pentru Business*, Editura ASE, Bucureşti, 2005.

Voineagu V., Ţiţan E., *Anchete şi sondaje*, volumele 1 şi 2, Editura „Andrei Şaguna", Constanţa, 2004.

*** *Anuarul Statistic al României*, INS, 1996-2005.

*** *Raportul dezvoltării umane în România*, INS, Bucureşti, 2000.

*** *Rezultatele recensământului populaţiei şi locuinţelor din martie 2002*, INS.

*** *Starea socială şi economică a României în anul 2004*, INS, Bucureşti, 2005.

*** www.insse.ro

*** http://europa.eu.int/comm/dgs/employment_social/index.html

Scanned with OKEN Scanner

www.meteorpress.ro

Statistica este un „numitor comun" al tuturor ştiinţelor, un limbaj comun al acestora; este un set de instrumente aplicabil în toate domeniile investigaţiei ştiinţifice, dar şi arta mânuirii corecte a acestora. Învăţându-ne să raţionăm corect, statistica ne ajută, astfel, să luăm deciziile optime în orice aspect al realităţii înconjurătoare.

Lucrarea de faţă îşi propune nu numai să treacă în revistă, dar şi să aprofundeze într-un mod complet, logic, unitar şi accesibil, bogăţia metodologiei statistice; ea se adresează studenţilor din învăţământul superior economic, dar şi celor din alte specializări ale învăţământului universitar, cercetătorilor, specialiştilor sau tuturor celor interesaţi în aplicarea vastei metodologii pe care statistica o pune la dispoziţie pentru întreaga cercetare ştiinţifică.

Distribuţie la:
Telefon/Fax: (021) 222.33.12
(021) 222.83.80
E-mail: editura@meteorpress.ro
carte@meteorpress.ro

ISBN 973-728-152-7
ISBN 978-973-728-152-4
9 789737 2815
28,50

Scanned with OKEN Scanner

O IDEE SCLIPITOARE- O LUCRARE ORIGINALĂ

„Vechi soiuri româneşti de viţă de vie''„„„„

A fost reeditată şi publicată în anul 2010 la Editura INTACT Bucureşti, ediţia a doua a lucrării „ *Vechi soiuri româneşti de viţă de vie'', elaborată de Dr. Ion M. Puşcă,* la un interval de numai patru ani de la prima apariţie.

Ideile originale care prezintă interes prin posibile rezultate pozitive, pentru care se întrevede un viitor spre aplicarea lor, au mobilizat activitatea în domeniile la care s-au adresat, contribuind la progres, în timp, după modul în care au fost adoptate. Esenţa lucrării Dr-lui Puşcă este prezentarea în sinteză a unei istorii milenare a soiurilor de viţăde vie cultivată pe Intinsele zone ale pământului nostru strămoşesc.Ea deşteaptă sentimentul patriotic, când deschidem prima copertă şi ne apare Dacia din Timpul lui Burebista(82-44 î.Cr.), ilustrând spaţiul despre care anticii Herodot(484-426 î. Cr.) şi Strabon(63-17 d.Cr.) ne-au lăsat scrierile lor cu referiri la existenţa viţei de vie şi consumului de vin la geto-daci.

La începutul cărţii, autorul face o confesiune profesional-sentimentală,pentru justiificarea preocupării şi efortului său de a elabora această lucrare, din care vedem calda sa afecţiune pentru profesia căreia şi-a dăruit munca de o viaţă şi dorinţa de a lăsa urmaşilor o zestre de informaţie istorică.

Prima parte oferă cititorului de azi, şi mai mult pentru generaţiile ce vor veni, o analiză profundă a relaţiei documentate dintre vin şi continuitatea poporului român, în spaţiul care au trăit strămoşii noştri. Este o documentare ce convinge asupra adevăruluiacestei relaţii, prin descrieri din cele mai vechi timpuri,care constituie ddovezi clare ; ele înscriu acest conţinut în istoria culturii noastre naţionale, cu fotţă de mobilizare a sentimentului preţuirii darurilor rămase de la înaintaşi.

Structurarea conţinutului lucrării, care cuprinde 246 pagini, este destinată pentru *Vechi soiuri româneşti de struguri pentru vin, l39 pagini ,Vechi soiuri româneşti de struguri de masă, 10 pagini, şi Vechi soiuri româneşti puţin cunoscute,lo panigi. In anexă mai aflăm 25 imagini color cu soiuri reprezentate prin forma strugurelui şi frunzei ; numeroase facsimile din presa vremii cu privire la prima ediţie a lucrării.*In această viziune, autorul a reuşit să prezinte o sinteză completă asupra vechiului cultivar de viţă de vie românească.Volumul se înceie *cu două pagini în care se prezintă un* Curriculum vitae , ce ne prezintă pe unul deintre iluştrii reprezentanţi ai oenologiei românşti, Dr. Ing. Ec. Ion M. Puşcă.

Autorul foloseşte o schemă originală pentru descrierea soiului, cu deosebiri interesante faţă de redarea în ampelorgrafiile clasice, oferind numeroase amănunte. Găsim unistoric detaliat ce convinge asupra originii certe a soiurilor în diferitelele regiuni viticole din ţară . Foarte atent descrieoriginea denumirii soiului primite de la strămunii noştri. Apoi, obiectivele clasice de caracterizare a soiului sunt respectate cu acurateţe. Dintre acestea reţinem, numărul suficient de însuşiri morfologice şi descrierea calităţii vinului, care permit recunoaşterea unui soi ; lucrarea fiind utilizabilă şi ca bibliografie didactică.

Peste toate cele arătate, citirea acestei lucrărieste un valoros act de cultură, produce o caldă satisfacţie, un motiv de admiraţie şi respect pentru trecutul nostru istoric în domeniul viei şi vinului ; o întărire a sentimentului de preţuire a zestrei lăsate de făuritorii neamului şi a patriei noastre.

Nu în ultimul rând, trebuie să elogiem stilul redării lucrării, într-un limbaj ştiinţific corect, o exprimare curată şi clară, fără nici o tendinţă de emfază stilistică,foarte agrebilă şi plină de învăţătură pentru cititor ; o scriere românească curată şi frumoasă.

Scanned with OKEN Scanner

Cartea Dr-lui Ion M. Puşcă se înscrie în şirul istoric al celor mai importante lucrăro cu fond ampelografic şi ampelkografii din istoria ştiinţei viticole româneşti., încheind în anul 2010 un şir de mărgăritate pentru bibliografia românească a domeniului. Prin această lucrare numele său se înscrie între cei mai de seamă autori ai acestui gen.

Pentru importanţa şi valoarea lucrării o vom aşeza în şirul celor mai reprezentattive cărţi şi elaboraări cu semniificaţie istorică în apariţia lor : *In anul l874, se publică în suplimentul Economia naţiunală al Revistei ştiinţifice, la rubrica – Încercări ampelografice- articolul ,, Speţele de struguri ce se găsesc în judeţul Prahova'', care este prima lucare in istoria genului, unde găsim tot prntru ptima oară trmenul ştiinţific – ampelografie sub semnătura P.S. Aurelianu, acela care a rămas cel mai important agronom şi economist al secolului XIX, la români. Cronologic a urmat prima ampelografie rpmâneaască elaborată de ilistrul agronom George Nicoleanu, sub titlul,, Introduction a l'ampelographie roumaine'',1900 ; Siuri de viţă de vie ,din primul ,, Tratat de viticultură'' a lui Vasile Brezeanu, fondator al învăţământului viticol la Herăstrău , 1902 ;,, Cele mai recomandate specii şi varietăţi mde viţă de vie'', de I.C. Teodorescu, l93o , umaşul lui Vasile Brezeanu la catesra de viticultură ;,,Ampelografia şi selecţia viţei de vie'' de Teodor Martin şi Mrin Neagu, l953,manual pentru şcoli ntehnice, a doua lucrare de gen ; ,, Ampelografie'' vol I-VIII,distinsă cu premii internaţionale, sub coordonarea acad. Gherasim Constantinescu,1959- l97o ; ,, Ampelografie'' vol I_II. De C. Ţârdea şiLiliana Rotaru, 2oo3. Incheiem această salbă de opere cu ,,* Vechi soiuri româneşti de viţă de vie'' de I.M. Puşcă. 2oo6, 2010.

Meritul autorului este preţuit de cunoscători şi de iubgitorii vinului, cum putem vedea în facsimilele din lucrare, dar va fi admirat şi păstrat în memoria cititorului obişnuit, iubitor de ţară şi interesat de cunoaşterea trecutului, care întotdeauna este o armă pentru construirea viitorului spre progres. Aducem calde aprecieri autorului pentru efortul şi calităţile deosebite de specialist şi patriot luminat, ouse în ace tă lucrare ce va dăinui peste secole.

Prof.dr. Viorel Peştean
Iaşi 6 marie 20ll

Scanned with OKEN Scanner